JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Docking-Based Comparative Intermolecular Contacts Analysis as New 3-D QSAR Concept for Validating Docking Studies and in Silico Screening: NMT and GP Inhibitors as Case Studies

Mutasem O. Taha,[†,*] Maha Habash,[‡] Zeina Al-Hadidi,[†] Amal Al-Bakri,[§] Khaled Younis,[∥] and Suhaib Sisan[∥]

[†]Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan

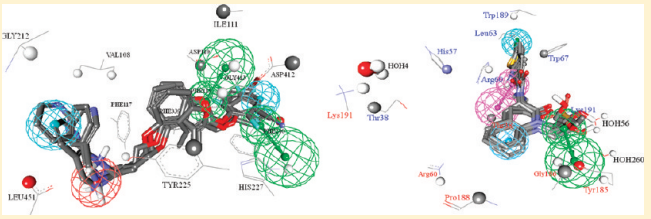
[‡]Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science University, Amman, Jordan

[§]Department of Pharmaceutics and Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan

[∥]Department of Computer Engineering, Faculty of Engineering, University of Jordan, Amman, Jordan

Supporting Information

ABSTRACT: The significant role played by docking algorithms in drug discovery combined with their serious pitfalls prompted us to envisage a novel concept for validating docking solutions, namely, docking-based comparative intermolecular contacts analysis (dbCICA). This novel approach is based on the number and quality of contacts between docked ligands and amino acid residues within the binding pocket. It assesses a particular docking configuration on the basis of its ability to align a set of ligands within a corresponding binding pocket in such a way that potent ligands come into contact with binding site spots distinct from those approached by low-affinity ligands and vice versa. In other words, dbCICA evaluates the consistency of docking by assessing the correlation between ligands' affinities and their contacts with binding site spots. Optimal dbCICA models can be translated into valid pharmacophore models that can be used as 3-D search queries to mine structural databases for new bioactive compounds. dbCICA was implemented to search for new inhibitors of candida N-myristoyl transferase as potential antifungal agents and glycogen phosphorylase (GP) inhibitors as potential antidiabetic agents. The process culminated in five selective micromolar antifungal leads and nine GP inhibitory leads.



1. INTRODUCTION

Structure-based drug design is considered one of the most important tools in drug discovery.[1] It can be divided into two major methodologies: de novo and docking-based design. De novo design involves the use of computer algorithms that construct virtual ligands inside the binding pocket.[1−3] On the other hand, docking involves fitting virtual ligands, usually from large virtual libraries, into the binding site employing algorithms that rely on force fields to calculate attractive and repulsive interactions between complementary groups within virtual ligand−protein complexes.[1−4]

Docking requires the availability of a three-dimensional structure for the targeted receptor or enzyme, which is generally obtained via X-ray crystallographic scattering, nuclear magnetic resonance (NMR), or homology modeling.[1−4] However, reliance on crystallographic structures represents a major problem for docking. Crystallographic structures are limited by inadequate resolution[9] and crystallization-related artifacts of the ligand−protein complex.[10−12] Moreover, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substrates.[13]

Moreover, molecular docking, which is basically a conformational sampling procedure in which various docked conformations are explored to identify the right one, can be a very challenging problem given the degree of conformational flexibility at the ligand−macromolecular level.[14−16] Docking programs employ diverse methodologies to evaluate different ligand conformations within binding pockets.[17−28]

Conformational sampling must be guided by a scoring function to evaluate the fitness between the protein and the ligand.[4,29−34] The final docked conformations are also selected according to their scores. The accuracy of the scoring function has a major impact on the quality of molecular docking results.[8,29,30,35,36] Scoring functions can be roughly grouped into three categories: force field methods,[19,20,23,37] empirical scoring functions,[18,38−42] and knowledge-based potentials.[43−47]

Nevertheless, although modern docking methods are able to calculate fairly accurately the position and orientation of a potential ligand in a receptor binding site, their major problem is the inability of scoring functions to evaluate binding free energies correctly in order to rank different potential ligand−receptor complexes.[1,3,6,8,10,85] The main problem in affinity prediction is that the underlying molecular interactions are highly complex, and various terms should be considered to quantify the free energy of the interaction process.[5−8,10,48,49]

Received: September 15, 2010
Published: March 03, 2011

Accordingly, the molecular modeler must find the optimal combination of docking−scoring algorithms capable of correct ranking of docked conformers and poses for potential ligands within certain binding pocket.

In addition to deciding the optimal docking−scoring combination for a particular docking problem, the molecular modeler must decide whether to leave crystallographically explicit water molecules in the binding site or not prior to ligand docking.[50−55] Furthermore, the fact that crystallographic structures lack information on hydrogen atoms means that it should be appropriately assumed, prior to docking, whether the ligand's ionizable moieties embedded within the binding site exist in their ionized form or not.[54,56] Reliance on pK_a values can be misleading because ligand ionizabilities depend on their local microenvironments within the binding pockets.[56]

These problems render docking experiments unreliable and make it imperative to adopt some measures to validate the docked conformers and poses for subsequent structure-based discovery or design.[84] Currently, these validation methods can be classified as follows. (i) Self-docking:[57] Cocrystallized ligands are removed from their corresponding binding sites and redocked employing the docking configuration under evaluation. The docking trial is generally considered successful if it can reproduce the cocrystallized conformation/pose of the ligands.[85] Unfortunately, this approach ignores the fact that docking experiments are usually performed to dock ligands into binding pockets patterned (imprinted) by other cocrystallized ligand(s), i.e., cross-docking, which reduces the validity of self-docking as validation technique in cross-docking situations.[57] (ii) Testing the ability of particular docking configuration to classify compounds in structural databases into actives and inactive: Normally, such databases are large lists of decoys seeded with a limited number of known active compounds.[78] The results are usually presented using receiver−operating characteristic (ROC) curves. Nevertheless, this approach suffers from a major drawback: it assumes inactivity of decoy molecules despite lack of supporting evidence.[58] (iii) Validation through 3-D QSAR methods: In this case, a particular docking configuration is considered valid if it succeeds in aligning a set of known ligands (i.e., into the binding pocket) in a 3-D alignment capable of explaining bioactivity variation, e.g., via CoMFA.[7,59,60] However, this approach is quite time-consuming and laborious as CoMFA grids include thousands of energy field points that require energy calculations and subsequent statistical filtering to search for field points that can explain bioactivity variation.

Pitfalls of structure-based design combined with inadequacies of corresponding validation methods prompted us to envisage a novel 3-D QSAR analysis to validate docking configurations: docking-based comparative intermolecular contacts analysis (dbCICA). This approach is based on the number and quality of contacts between docked ligands and amino acid residues within the binding pocket. dbCICA assesses a certain docking configuration based on its ability to align a set of ligands (i.e., within a corresponding binding pocket) in such a way that potent ligands come into contact with binding site spots distinct from those approached by low-affinity ligands and vice versa. In other words, dbCICA evaluates the consistency of docking by assessing the correlation between ligands' affinities and their contacts with binding site spots. To test this new technique, we carried out dbCICA modeling against two interesting targets: candida N-myristoyl transferase (CaNMT) and glycogen phosphorylase (GP). The resulting models were used as 3-D search queries to catch new hits that were subsequently tested against each corresponding target.

The significance of candida N-myristoyl transferase (CaNMT) as a target in antifungal research[64−68,86−88] combined with the availability of appropriate crystallographic structures for this target[89−91] prompted us to evaluate dbCICA against this target. Toward this end, a diverse list of 108 CaNMT inhibitors were docked into the binding pocket of CaNMT employing diverse docking conditions that included three docking engines, 14 scoring functions, two ligand ionization states (ionized versus un-ionized), and two binding site hydration states (hydrous and anhydrous).

Similarly, the significance of glycogen phosphorylase (GP) as valid antidiabetic target,[93−95] combined with the availability of several human liver GPa X-ray complexes,[96−107] prompted us to evaluate a variety of docking−scoring conditions to identify optimal parameters capable of aligning GP inhibitors within GP binding site in such a way to be consistent with their bioactivities. A list of 35 GP inhibitors were docked into the binding pocket of GP employing ligandfit docking engine via six scoring functions, two ligand ionization states (ionized versus un-ionized), and two binding site hydration states (hydrous and anhydrous).

Subsequently, dbCICA modeling was employed in both examples as a gauge to measure the success of each set of docking parameters. Thereafter, optimal dbCICA models were employed as templates to generate pharmacophore models that were employed as search queries to screen the National Cancer Institute (NCI) structural database for new GP and CaNMT inhibitors. Hits were subsequently bioassayed. Several hits illustrated good biological properties.

2. RESULTS AND DISCUSSION

2.1. Basic Concept of dbCICA. Although docking engines suffer from being unable to calculate free energy of binding, they normally succeed in reproducing cocrystallized ligand poses and conformations among their high-ranking docking solutions.[1,3,6,8,10,85] This suggests that it is quite possible to correlate docked 3-D ligand poses and conformers with bioactivities, as shown previously using CoMFA modeling.[7,59,60]

In dbCICA, the interest is focused on identifying a set of atoms within the binding site that tend to contact with potent docked ligands, while avoid poorly active docked ligands. If such a set of contact atoms is identified for a docked list of ligands, then one can assume that the docking configuration is successful, i.e., it arranged the molecules in a such way that can explains variation in bioactivity.

Undoubtedly, high ligand−receptor affinity is mediated by a certain critical number of interactions, i.e., potent ligands usually satisfy certain critical interactions (hotspots) within the binding pocket. However, the fact that docking engines and scoring functions evaluate a large number of ligand−receptor interactions to generate their docking solutions means the influence of critical interactions might be overlooked. In this context, identifying a set of affinity-discriminating contact atoms within the binding site should help not only to validate a particular docking configuration, but also in pinpointing the critical ligand−receptors interactions responsible for affinity as such discriminatory contacts encode for nearby attractive interactions. In fact, the resulting dbCICA models can be translated into pharmacophoric models of a limited number of critical binding features that can be used as 3-D search queries to mine for new ligands.

2.2. Practical Aspects and Implementation of dbCICA. The dbCICA concept leaves a question about the definition of "contacts", i.e., the interatomic distance thresholds that can be considered as acceptable contacts. The fact that most ligand−receptor interactions, including hydrogen bonding and van der Waals' forces, illustrate optimal strength at distance ranges of 2.0−3.5 Å prompted us to perform dbCICA analysis by implementing three distance thresholds as interatomic contacts, i.e., 2.0, 2.5, or 3.5 Å. Interatomic distances below (or equal to) the predefined threshold are considered as contacts and are encoded by a binary value of one, while larger distances are considered as noncontacts and encoded by zeros. Interatomic distances are calculated on the basis of explicit hydrogen atoms.

As mentioned earlier, discriminatory ligand−receptor contacts are surrogates of nearby critical ligand−receptor interactions. However, because ligand−receptor affinity is normally mediated by a set of concomitant critical attractive interactions, it is expected that their proxy contacts are also concomitant. This basic principle is represented in dbCICA analysis by searching for discriminatory ligand−receptor contacts that have their summation values directly proportional to bioactivity, i.e., search of concurrent contacts rather than separate contacts.

Furthermore, our dbCICA algorithm has the option of allowing any particular positive contacts column to emerge up to three times in the optimal summation model, i.e., allows variable weights for contacts. This option identifies intermolecular contacts of higher weights or contributions in the optimal dbCICA models. See point (iii) in docking-based comparative molecular contacts analysis (dbCICA) (section 4.1.6) of the Experimental Section for more details.

Still, some contacts are expected to be inversely proportional to bioactivity (i.e., negative contacts). These contacts encode for repulsive interactions (steric clashes). However, to remove any correlation faults resulting from summations of negative and positive contacts during the search for optimal contact combinations, dbCICA analysis is preceded by scanning the correlations between each contact column and bioactivity. Inversely proportional contacts, i.e., negative contacts, are removed during the initial search phase for optimal combinations of positive contacts.

Nevertheless, because ligand−receptor binding is controlled by both attractive and repulsive forces, both positive and negative contacts are later combined in dbCICA models. However, to maintain the consistency of correlation calculations, i.e., maintain the trend of direct proportionality between contacts combinations and bioactivities, it was decided to invert negative contacts, i.e., by converting their zeros to ones and vice versa. Subsequently, a second search phase is performed to find optimal summations of negative contacts that upon combination with previously defined optimal positive contacts summations yield optimal correlations with bioactivity.

It remains to be mentioned that dbCICA contacts summation models are judged on the basis of three success criteria: correlation coefficient (r^2), leave-one-out r^2 (r^2_{LOO}), or K-fold r^2 ($r^2_{K\text{-fold}}$). See genetic algorithm implementation in dbCICA modeling (section 4.1.7) of the Experimental Section for more details.

2.3. Inferences from dbCICA Output. From the previous discussion, one can summarize information collected from dbCICA modeling in the following points.

- Success in identifying a combination of ligand−receptor contact points of collinear relationship with bioactivity suggests validity of the corresponding docking−scoring approach. This is reminiscent of the use of other 3-D QSAR methods, e.g, CoMFA, to validate docking−scoring methodologies.[7,59,60]
- Successful dbCICA models point to critical amino acids involved in ligand binding (binding hot spots). This information combined with identifying the optimal docking−scoring conditions are of great help in subsequent structure-based lead optimization efforts.[27]
- Successful dbCICA models can be translated into pharmacophoric hypothesis useful for in silico screening of virtual databases to identify new hits. Although many studies have been published claiming to extract a pharmacophore starting solely from the structural analysis of one or few protein−ligand complexes, the extracted information cannot be termed "pharmacophore" if it is not based on proper structure−activity relationship analysis.[63] Although protein−ligand complexes can clear up some ambiguities about a binding pharmacophore, such complexes only display active molecules bound to the target site; they do not bring any information regarding the requirement of each interaction to the activity.[27] This makes dbCICA handy for building pharmacophore models as it is based on 3-D SAR differences between active and inactive ligands within the binding site.

2.4. First Case Study on dbCICA: Discovery of CaNMT Inhibitors. NMT is a cytosolic monomeric enzyme present in eukaryotes such as protozoa, fungi, and animals but absent in bacteria.[86,87] NMT catalyzes the myristoylation of protein targets by transfer of myristate from myristoyl-CoA to substrate protein. In fungi, myristoylation is involved in the regulation of interaction between certain signaling proteins and the cellular membrane, which is required in signal transduction cascades and vesicular and protein trafficking processes.[87,88] The main focus of recent efforts toward the development of new candidal NMT (CaNMT) inhibitors concentrate on structure-based ligand design.[88]

Accordingly, we decided to employ dbCICA to identify the optimal conditions and parameters necessary for successful docking of 108 CaNMT inhibitors (Table A of the Supporting Information). Furthermore, the optimal dbCICA model was converted into a pharmacophore hypothesis, which we used as a 3-D search query to mine for new anti-NMT hits from NCI list of compounds.

The study was performed as follows. The collected compounds were docked into the active site of CaNMT employing three docking engines: LigandFit,[25,70] LibDock,[27,28] and FRED.[26] In each case, four docking variations were implemented: two related to ligand ionization (ionized vs un-ionized ligands) and two related to binding site hydration (hydrous vs anhydrous binding site).

Subsequently, high-ranking docking solutions were scored by 14 different scoring functions. Seven scoring functions were imbedded in FRED and were therefore used for scoring FRED-generated docked poses and conformers, and the remaining seven were used to score the docked results of LigandFit and LibDock. The highest-ranking ligand conformers and poses, on the basis of each scoring function, were aligned to construct corresponding dbCICA models.

Intermolecular ligand binding site contacts were determined employing three distance thresholds: 2.0, 2.5, and 3.5 Å. Therefore, three corresponding contact binary matrices were generated for each docking solution (see section 4.1.6 of the Experimental

Table 1. Highest Ranking Anti-CaNMT dbCICA Models, Their Corresponding Parameters, and Statistical Criteria[a]

dbCICA model	docking engine	ligand ionization state	explicit water[b]	scoring function	contacts distance threshold[c]	number of positive contacts[d]	number of negative contacts[e]	r_{108}[f]	r^2_{LOO}[g]	$r^2_{5\text{-fold}}$[h]	F statistic
A-I	LibDock	ionized	absent	PMF04	3.5	7	5	0.73	0.51	0.51	119.5
A-II	LibDock	ionized	absent	PMF04	3.5	8	10	0.72	0.51	0.50	115.3
A-III	LibDock	unionized	present	PLP1	2.5	7	10	0.69	0.47	0.47	99.0
A-IV	LibDock	unionized	absent	PMF	3.5	6	10	0.69	0.46	0.46	96.2
A-V	FRED	ionized	present	CHEMSCORE	2.5	10	10	0.68	0.45	0.45	91.9
A-VI	LibDock	ionized	present	PMF	3.5	8	5	0.67	0.43	0.45	87.2

[a] The presented information in this table is extracted from Tables B, C, and D of the Supporting Information except for the F statistic, which is calculated from the correlation connecting $-\log(IC_{50})$ and the contacts sums of the corresponding dbCICA models. [b] Crystallographically explicit water of hydration. [c] Distance thresholds used to define ligand binding site contacts. [d] Optimal number of combined (i.e., summed) bioactivity-enhancing ligand binding site contacts. [e] Optimal number of combined (i.e., summed) bioactivity-disfavoring ligand binding site contacts. [f] Noncross-validated correlation coefficient for 108 training compounds. [g] Cross-validation correlation coefficients determined by the leave-one-out technique. [h] Cross-validation correlation coefficients determined by the leave-20%-out technique repeated 5 times.

Section). Subsequently, a genetic algorithm (GA)-based search was implemented to search for the best combination (summation) of ligand−receptor intermolecular contacts capable of explaining bioactivity variation across the training compounds. GA was instructed to scan combinations of 2−10 directly proportional intermolecular (positive) contacts followed by 5 or 10 inversely proportional (negative) contacts.

Tables B, C, and D of the Supporting Information show the contacts thresholds in angstroms, number of positive and negative contacts, and statistical criteria of the best dbCICA models based on docked solutions generated by FRED, LibDock, and LigandFit, respectively, while Table 1 summarizes the highest-ranking results of the three tables. Clearly, from Table 1, all high-ranking dbCICA models shared mediocre statistical criteria with $r^2_{5\text{-fold}}$ values around 0.50. We believe this behavior is attributed to the fact that in dbCICA modeling the bioactivities of training compounds are correlated with contacts summation, i.e., a single descriptor in each round of genetic iterations. Correlation with a single descriptor is expected to yield mediocre correlation statistics. However, correlation to multiple contacts failed to achieve any substantial prediction. Moreover, the fundamental theory of dbCICA is based on the occurrence of simultaneous contacts, which are best represented by summation. Therefore, we believe the best use of dbCICA statistical criteria is to merely rank dbCICA models and to prioritize them for translation into pharmacophore models.

Table 1 shows that LibDock illustrated the best results with five of its docking settings yielding dbCICA models of $r^2_{5\text{-fold}}$ values ≥ 0.45. FRED came next with only one dbCICA model of $r^2_{5\text{-fold}} \geq 0.45$. Interestingly, PMF-based scoring was superior over other scoring functions as four out of the six of the best dbCICA models were based on PMF. Similarly, ligand ionization seems to enhance the qualities of the dbCICA models as four out of the six best dbCICA models were based on docking solutions performed on ionized ligands. However, the significance of explicit water molecules seems to be less drastic as three out of the six successful dbCICA models were based on hydrous binding sites, albeit they were statistically better models (i.e., dbCICA models **A-I** and **A-II**, Table 1).

The fact that both dbCICA models **A-I** and **A-II** (Table 1), which shared the best statistical success criteria, were based on the same docking−scoring conditions, i.e., LibDock/PMF04-based docking of ionized ligands into anhydrous binding pocket, greatly supported the validity of their docking conditions and

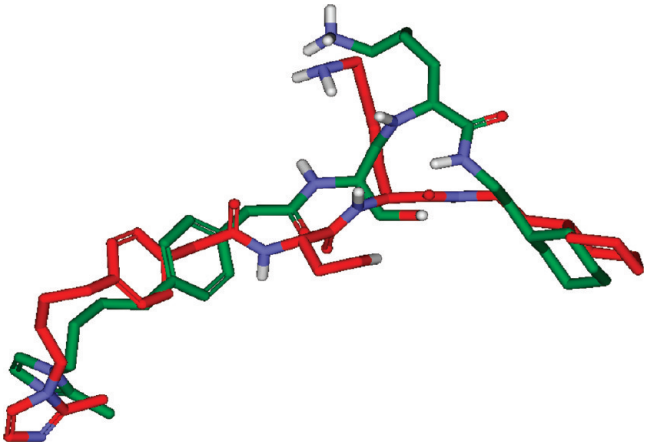


Figure 1. Comparison between the cocrystallized pose of MIM (red, PDB code: 1IYK, resolution 2.3 Å) and its docked pose into CaNMT according to docking conditions of models **A-I** and **A-II** (green, Table 1).

therefore provided excellent frameworks for subsequent development into promising pharmacophoric hypotheses for in silico screening. Figure 1 compares how these conditions dock a ligand compared to its cocrystallized structure.

Table 2 shows critical CaNMT binding site contact atoms and their weights as proposed by optimal dbCICA models **A-I** and **A-II**. Clearly from Table 2, a number of critical contacts emerged in both dbCICA models **A-I** and **A-II**, which further support their significance. For example, the carbonyl oxygens of LEU451 and PHE240 emerged in both models pointing to their exceptional significance in controlling CaNMT bioactivity.

Figures 2 and 4 show how dbCICA models **A-I** and **A-II** were translated into corresponding pharmacophoric models (**hypoA-I** and **hypoA-II**, respectively, Figures 3 and 5) employing DS 2.5 environment. See molecular field analysis (section 4.1.8) of the Experimental Section for more details. Initially, the binding pockets were annotated by rendering significant contacts atoms in spherical forms (see Figures 2A and 4A). Subsequently, we selected few potent ($IC_{50} < 0.05\ \mu M$) and well-behaved docked compounds, i.e., of least difference between experimental and fitted bioactivities (determined by regressing the inhibitors' bioactivities and their docking-based contacts summations). Thereafter, appropriate pharmacophoric features were placed

dx.doi.org/10.1021/ci100368t |J. Chem. Inf. Model. 2011, 51, 647–669

Table 2. Critical Binding Site Contact Atoms Proposed by Optimal dbCICA Models for CaNMT Inhibitors

dbCICA model	favored contact atoms (positive contacts)[a,b]: amino acids and corresponding atom identities[c]	weights[d]	disfavored contact atoms (negative contacts)[e]
A-I	ILE111: CA	2	ASP112:OD1; GLY212:O; HIS227:HD1; PHE117:HE1; TYR335:CZ
	LEU451: O	3	
	PHE115: HB1	1	
	PHE240: O	3	
	SER214: HN	3	
	TYR225: CD1	1	
	TYR335: HH	1	
A-II	ASP110: CB	3	ASP412:CB; GLY212:HA1; GLY413:HA1; GLY413:HA2; ILE111:CG1; PHE115:CZ; PHE117:HE1; PHE339:HE1; TYR225:CD2; VAL108:HG11
	HIS227: HD2	3	
	ILE111: CB	1	
	LEU450: C	1	
	LEU451: O	3	
	PHE240: O	3	
	THR211: HG1	1	
	VAL108: HG21	3	

[a] Numbers and distance thresholds are as in Table 1. [b] Bioactivity-proportional ligand binding site contacts. [c] Binding site amino acids and their significant atomic contacts. Atom codes are as provided by the protein data bank file format (e.g., ILE111: CA encodes for carbon atom (A) of isoleucine number 111) except for hydrogen atoms which were coded by DS2.0. [d] Degree of significance (weight) of corresponding contact atom, which points to times it emerges in the final dbCICA model (see point (v) in section 4.1.6 of the Experimental Section). [e] Bioactivity-disfavoring ligand binding site contacts (steric clashes).

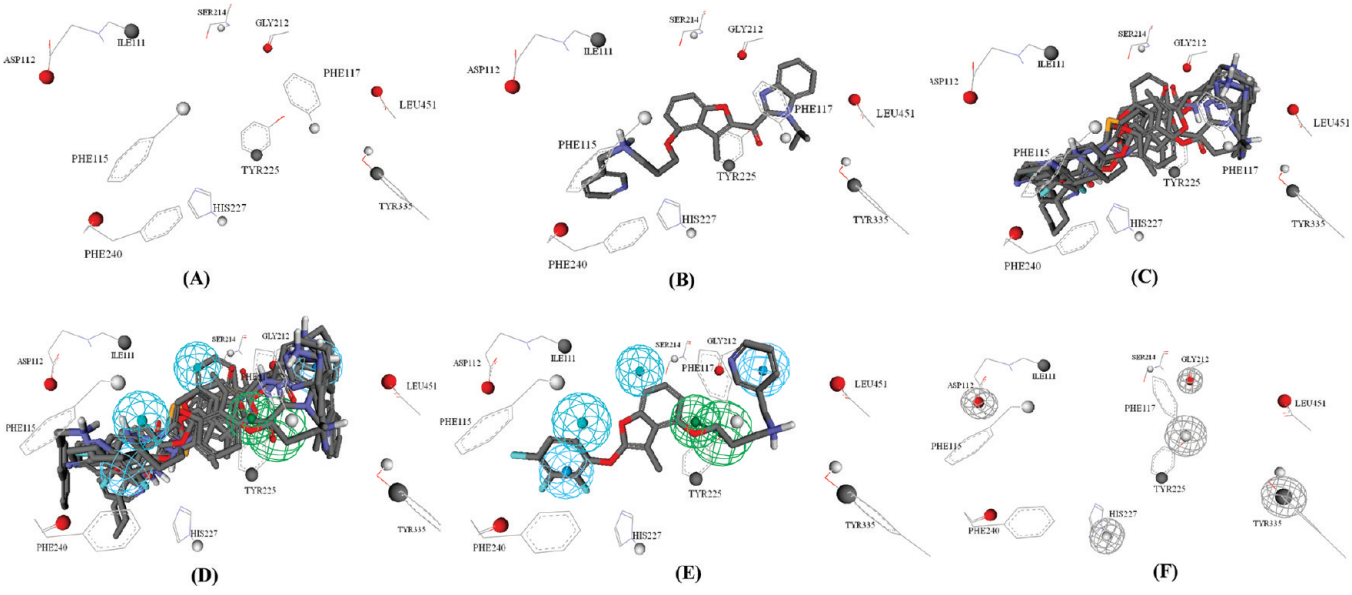


Figure 2. Steps for manual generation of binding hypothesis **HypoA-I** as guided by dbCICA model **A-I** (Tables 1 and 2). (A) The binding site moieties in dbCICA model **A-I** with significant contact atoms shown as spheres. (B) The docked pose of the well-behaved compound **A103** (IC_{50} = 0.0013 μM) within the binding pocket. (C) The docked poses of the well-behaved and potent compounds **A3**, **A24**, **A78**, and **A103**. (D) Manually placed pharmacophoric features onto chemical moieties common among docked well-behaved potent compounds **A3**, **A24**, **A78**, and **A103**. (E) The docked pose of **A78** (well-behaved and potent, IC_{50} = 0.0057 μM) and how it relates to the proposed pharmacophoric features. (F) Exclusion spheres fitted against binding site atoms showing negative correlations with bioactivity (as emergent in dbCICA model **A-I**).

onto common chemical functionalities among aligned docked compounds, as in Figures 2D and 4D.

The pharmacophoric features were placed in such away to highlight the interactions encoded by nearest critical contacts.

dx.doi.org/10.1021/ci100368t |J. Chem. Inf. Model. 2011, 51, 647–669

For example, in case of dbCICA model **A-II**, emergence of a significant contact at the carboxylic oxygen of LEU451, combined with the consensus of well-behaved, potent docked ligands on placing nearby quaternary ammonium groups, prompted us to place a positive ionizable feature onto the ammonium groups (Figure 4D). Similarly, agreement of docked, potent and well-behaved compounds on placing hydrophobic moieties (aromatic rings) near to hydrogen (HG21) and oxygen atoms of VAL108 and PHE240, respectively (both of which emerged as significant contacts in dbCICA model **A-II**, as shown in Table 2) prompted us to place hydrophobic pharmacophoric features onto both positions, as in Figure 4D. The significant dbCICA contact at PHE240 oxygen seems to encode for $\pi-\pi$ stacking of aromatic rings in docked ligands against the phenyl side chain of PHE240. Finally, agreement among docked potent and well-behaved training compounds in placing hydrogen bond accepting moieties directed toward β-carbon (CB) of ASP110 and HB1 hydrogen of HIS227, encouraged us to place two hydrogen-bond acceptor features originating from the benzofuran oxygens and aromatic fluoro (or cyano) substitutents of the docked ligands pointing toward the β-carbon of ASP110 and HB1 hydrogen of HIS227, respectively. Incidentally, the significant contact at the β-carbon of ASP110 seems to encode for hydrogen bonding involving the carboxylic acid of ASP110 and the benzofuran oxygens of the docked ligands.

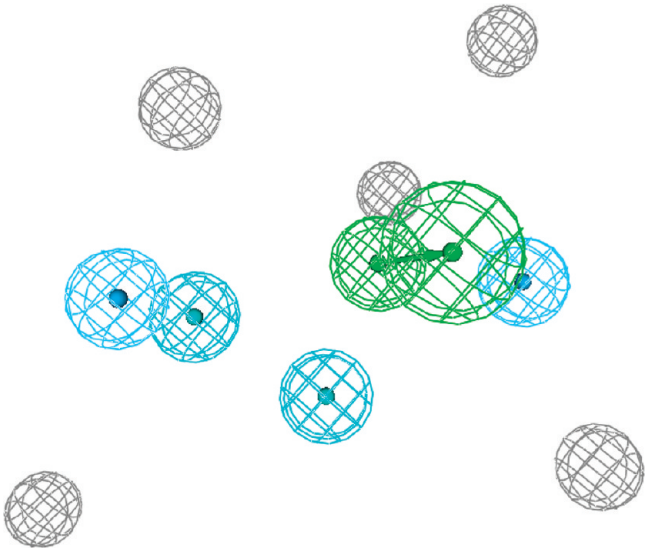


Figure 3. HypoA-I pharmacophoric features and exclusion spheres. Light blue spheres represent hydrophobic features, dark blue spheres represent hydrophobic aromatic features, vectored green spheres represent hydrogen bond acceptor, and gray spheres represent exclusion regions.

A similar strategy was implemented for the development of pharmacophore model **HypoA-I** starting with dbCICA model **A-I**. Figures 3 and 5 show the two pharmacophore models, including their exclusion regions, while Table E of the Supporting Information shows the *X*, *Y*, *Z* coordinates of the two pharmacophores.

2.4.1. Validation of Optimal Anti-CaNMT dbCICA Models. Three validation procedures were implemented to assess the dbCICA models, namely, CoMFA, self-docking, and ROC analyses.

(1). CoMFA Modeling. To validate optimal dbCICA models generated from CaMNT inhibitors (models **A-I** to **A-VI**, Table 1), it was decided to assess whether their corresponding

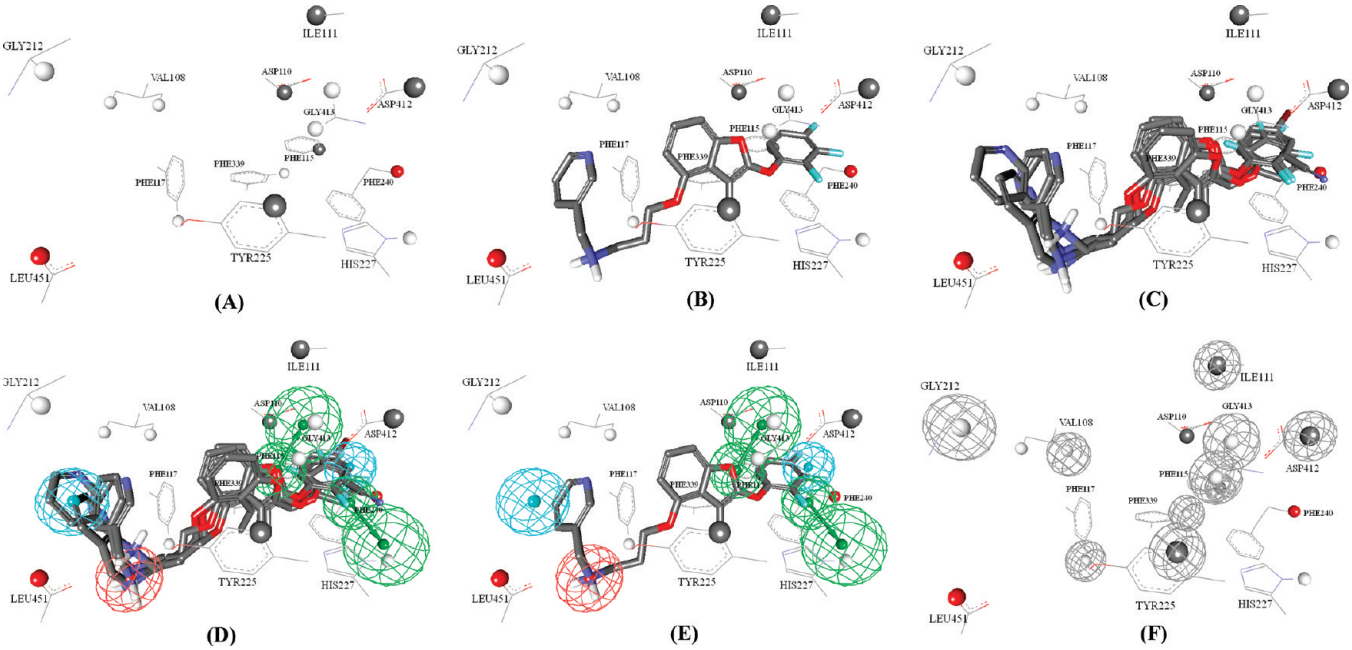


Figure 4. Steps for manual generation of binding hypothesis **HypoA-II** as guided by dbCICA model **A-II** (Tables 1 and 2). (A) The binding site moieties in dbCICA model **A-II** with significant contact atoms shown as spheres. (B) The docked pose of the well-behaved compound **A78** (IC_{50} = 0.0057 μM) within the binding pocket. (C) The docked poses of the well-behaved and potent compounds **A76**, **A78**, **A79**, and **A93**. (D) Manually placed pharmacophoric features onto commonly aligned chemical moieties among docked well-behaved potent compounds **A76**, **A78**, **A79**, and **A93**. (E) The docked pose of **A78** and how it relates to the proposed pharmacophoric features. (F) Exclusion spheres fitted against binding site atoms of negative correlations with bioactivity (as emergent in dbCICA model **A-II**).

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

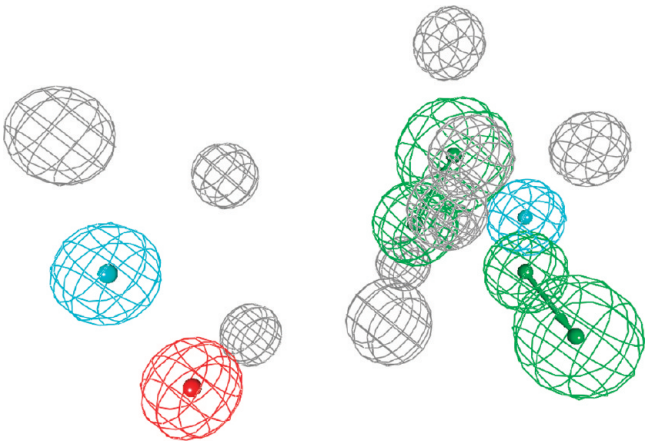


Figure 5. HypoA-II pharmacophoric features and exclusion spheres. Light blue spheres represent hydrophobic features, vectored green spheres represent hydrogen bond acceptors, red sphere represents positive ionizable feature, and gray spheres represent exclusion regions.

molecular alignments yield self-consistent and predictive CoMFA models.[7,59,60] For comparison purposes, a similar assessment was performed for molecular alignments of a low-ranking dbCICA model (e.g, LibDock-based docking of ionized ligands into anhydrous binding pocket employing PLP1 scoring function of $r^2_{5\text{-fold}} = 0.31$, Table C of the Supporting Information). Table 3 shows the statistical criteria of the resulting CoMFA models.

Clearly from Table 3, the statistical criteria of the top two dbCICA models (i.e., **A-I** and **A-II**, Table 2) correlate nicely with those of their CoMFA counterpart. The remaining models failed either in their r^2_{LOO} or r^2_{PRESS} (Table 3), which also correspond to the low $r^2_{5\text{-fold}}$ of their corresponding dbCICA counterparts ($\leq$0.47, Table 1). Unsurprisingly, the low-ranking dbCICA model coincided with the low quality CoMFA model (Table 3). Agreement between CoMFA and dbCICA provides additional evidence of the capacity of dbCICA modeling as a validation technique in docking studies.

(2). Self-Docking. However, to assess the similarity between ligand poses and conformers produced by the docking conditions of dbCICA models **A-I** and **A-II** (Table 1) and the crystallographic structure of a bound ligand, we compared the docked pose of the ligand MIM (employing docking conditions of **A-I** and **A-II**) with its cocrystallized pose (PDB code: 1IYK, resolution 2.3 Å). Figure 1 shows alignment of the two structures. Clearly from the figure the docking conditions succeeded in placing the pharmacophoric features of the ligand in comparable fashion with that of the cocrystallized pose (rms = 3.26 Å), i.e., the cyclohexyl side chain, imidazol ring, primary amine, and central aromatic and amidic moiety were all closely aligned in both cases adding further support to the docking configuration. The slight shifts in the positions of different moieties can be attributed to fact that docking was performed on anhydrous binding pocket compared crystallographic structure.

(3). Receiver Operating Characteristic Curve Analysis of Corresponding Pharmacophores (HypoA-I and HypoA-II, Figures 3 and 5). The two pharmacophores were subjected to receiver-operating curve (ROC) analysis. In ROC analysis, the ability of a particular pharmacophore model to correctly classify a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the corresponding ROC as well as other parameters: overall accuracy, overall specificity, overall true positive rate, and overall false negative rate (see section SM-2 of ther Supporting Information for more details).[62,78−81,112−115] Figure 6 and Table 4 show the ROC performances of our dbCICA-based pharmacophores.

It is clearly from both the table and figure that **HypoA-II** significantly outperformed **HypoA-I**, which is probably attributed to the absence of a positive ionizable feature in **HypoA-I**. The superior properties of **HypoA-II** prompted us to use it as 3-D search query against the NCI list of compounds to discover new anti-CaNMT hits.

2.4.2. In Silico Screening and in Vitro Validation. **HypoA-II** was employed as a 3-D search query to screen the national cancer institute list of compounds (NCI, includes 238,819 compounds). NCI hits were subsequently filtered by a molecular weight threshold of 550 Da in order to remove large, nondrug-like compounds. The resulting hits were docked into CaNMT protein (PDB code: 1IYK) employing the same docking conditions of dbCICA model **A-II** (i.e., LibDock docking of ionized ligands without protein hydration water and employing PMF04 scoring function, as in Table 1). The resulting docked poses were

Table 3. Statistical Results of the CoMFA Models Obtained via High-Ranking Anti-CaNMT dbCICA-Based Docking−Scoring Combinations (as in Table *1*)

				statistical criteria					
dbCICA models[a]		docking conditions[a]	scoring function[a]	terms[b]	PC[c]	r^{2d}	r^2_{LOO}[e]	r^2_{BS}[f]	r^2_{PRESS}[g]
highest[h] ranking	**A-I** and **A-II**[h]	Libdock-ionized ligands:anhydrous binding pocket	**PMF04**	15	4	0.76	0.61	0.75	0.49 (0.59)[j]
	A-V	Fred-ionized ligands: hydrous binding pocket	**CHEMSCORE**	17	3	0.77	0.61	0.76	0.36
	A-III	Libdock-un-ionized ligands: hydrous binding pocket	**PLP1**	17	3	0.70	0.30	0.47	0.55
	A-IV	Libdock-un-ionized ligands: anhydrous binding pocket	**PMF**	17	5	0.70	0.30	0.39	0.45
	A-VI	Libdock-ionized ligands: hydrous binding pocket	**PMF**	16	6	0.77	0.57	0.68	0.12
low ranking[i]		Libdock-ionized ligands: anhydrous binding pocket	**PLP1**	16	6	0.737	0.351	0.521	0.33

[a] dbCICA models and corresponding docking−scoring conditions were selected from Tables B and C of the Supporting Information. [b] Number of CoMFA descriptors in the best 3-D QSAR model. [c] Number of principal components (latent variables) in the best 3-D QSAR model. [d] Noncrossvalidated correlation coefficient for 87 training compounds. [e] Cross-validation correlation coefficients determined by the leave-one out technique. [f] Bootstrapping correlation coefficient. [g] Predictive r^2 determined for the 21 test compounds (see Experimental Section). [h] dbCICA models numbered as in Table 1. [i] Example of low-ranking dbCICA model selected from Table C of the Supporting Information. [j] r^2_{PRESS} values after removing single outlier from the testing set (**A85**, Table A of the Supporting Information).

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

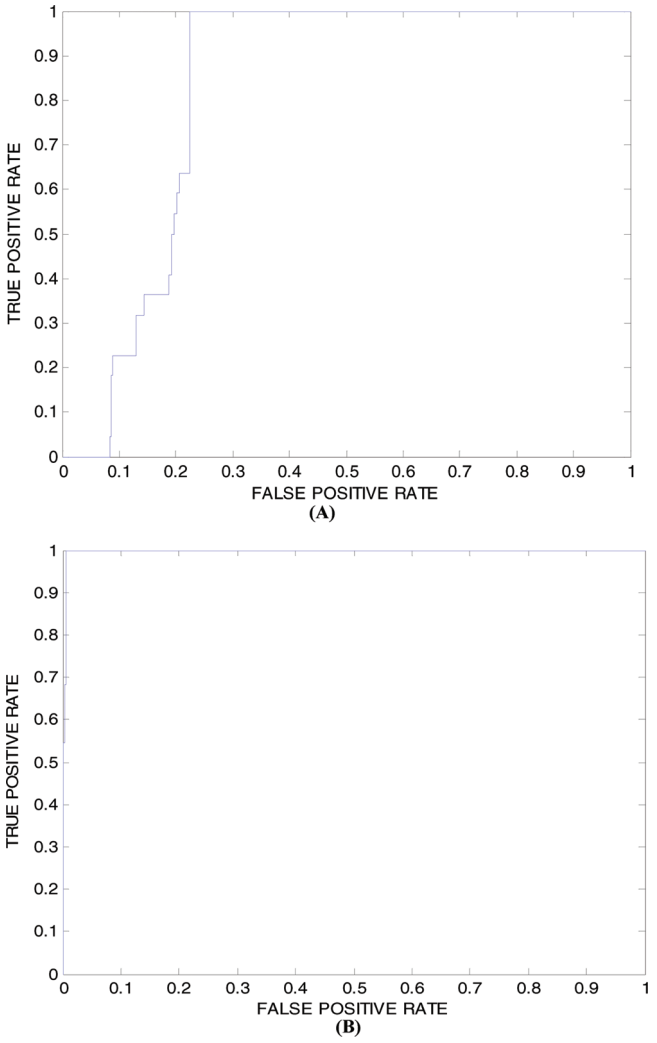


Figure 6. Receiver operating characteristic (ROC) curves of dbCICA-based pharmacophores (A) **HypoA-I** and (B) **HypoA-II**.

Table 4. Performance of dbCICA-Based Anti-CaNMT Pharmacophores as 3-D Search Queries

pharmacophoremodel	ROC[a]and AUC[b]	ACC[c]	SPC[d]	TPR[e]	FNR[f]
HypoA-I	82.70	96.41	97.63	63.64	23.73
HypoA-II	99.82	96.41	96.78	86.36	3.22

[a] ROC: receiver operating characteristic. [b] AUC: area under the curve (%). [c] ACC: overall accuracy (%). [d] SPC: overall specifity (%). [e] TPR: overall true positive rate. [f] FNR: overall false negative rate (%).

subsequently analyzed for critical contacts according to dbCICA model **A-II** (Table 2) and the sums of critical contacts for each hit compound were employed to rank the corresponding hits and prioritize subsequent in vitro testing. Table 5 and Figure 7 show the highest ranking hits and their experimental in vitro bioactivities.

Out of the the 35 highest-ranking hits acquired for experimental validation, 11 were found to possess potent antifungal minimal inhibitory concentrations (MICs) ranging from 46.7 to 855 μM. Five hits, namely, **A111** (NCI0380503), **A112** (NCI0009609), **A120** (NCI0057144), **A125** (NCI0111726),

Table 5. High-Ranking Hits Captured by HypoA-II, Their Corresponding Summation Contacts According to dbCICA model A-II, and in Vitro Bioactivities

hit[a]	NCI code	dbCICA contacts[b]	MIC (μM) against[c] *C. albicans*
A109	NCI0349841	20	NA[d]
A110	NCI0040089	12	NA
A111	NCI0380503	17	107
A112	NCI0009609	22	221
A113	NCI0035706	11	NA
A114	NCI0032470	10	NA
A116	NCI0062371	15	NA
A117	NCI0087942	13	NA
A118	NCI0036837	16	NA
A119	NCI0092064	15	NA
A120	NCI0057144	19	194
A122	NCI0296300	14	NA
A123	NCI0292266	18	NA
A124	NCI0291562	20	NA
A125	NCI0111726	20	373
A126	NCI0117460	17	NA
A127	NCI0119940	10	NA
A128	NCI0132389	15	NA
A129	NCI0051663	13	46.7
A130	NCI0318194	15	NA
A131	NCI0026853	20	NA
A132	NCI0278358	14	NA
A133	NCI0032457	12	NA
A134	NCI0683339	18	NA
A135	NCI0691085	15	NA
A136	NCI0270746	10	NA
A137	NCI0014500	17	NA
A138	NCI0064977	19	NA
A139	NCI0280900	18	NA
A140	NCI0327929	13	NA
	miconazole[e]	–	28

[a] Hits numbers as in Figure 7. [b] Contacts summations according to dbCICA model **A-II** (Table 2). [c] Minimum inhibitory concentration determined for each hit compound against *C. albicans*. [d] Not active (at concentrations ≤1 mg/mL). [e] Reference antifungal agent.

and **A 129**(NCI0051663) exhibited selective antifungal activities, i.e., they were inactive against gram positive or gram negative bacteria exemplified by *S. aureus* and *E. coli*, respectively. Figure 8 shows how a selective anti-NMT hit dock into the binding pocket of this enzyme and how the docked conformer maps **HypoA-II**. Needless to say, the selective antifungal bioactivities strongly suggest inhibition of NMT because this enzyme exists only in fungal cells, i.e., not in bacteria (neither gram positive nor gram negative).

2.5. Second Case Study: Discovery of GP Inhibitors. Glycogen phosphorylase (GP), a homodimeric protein,[108] catalyzes glycogenolysis by converting glycogen to glucose-1-phosphate in the liver.[94,108,109] Accordingly, there is great recent interest in the discovery of new GP inhibitors as possible treatments for diabetes.[110,111]

The study was performed as follows. Initially all 35 compounds (**B1**-**B35**, Table F of the Supporting Information) were

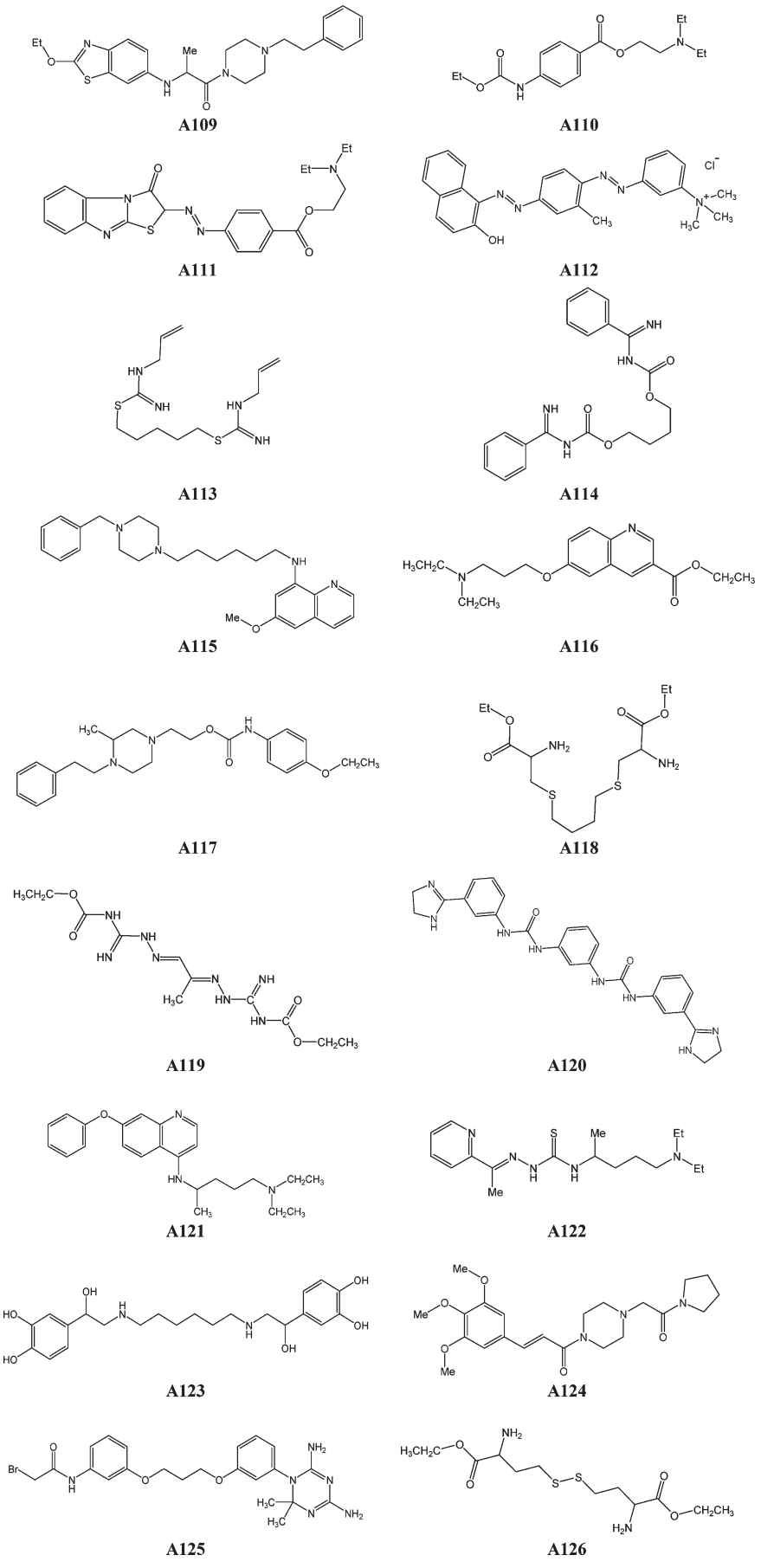


Figure 7. Chemical structures of the tested highest ranking hits captured by **HypoA-II**.

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

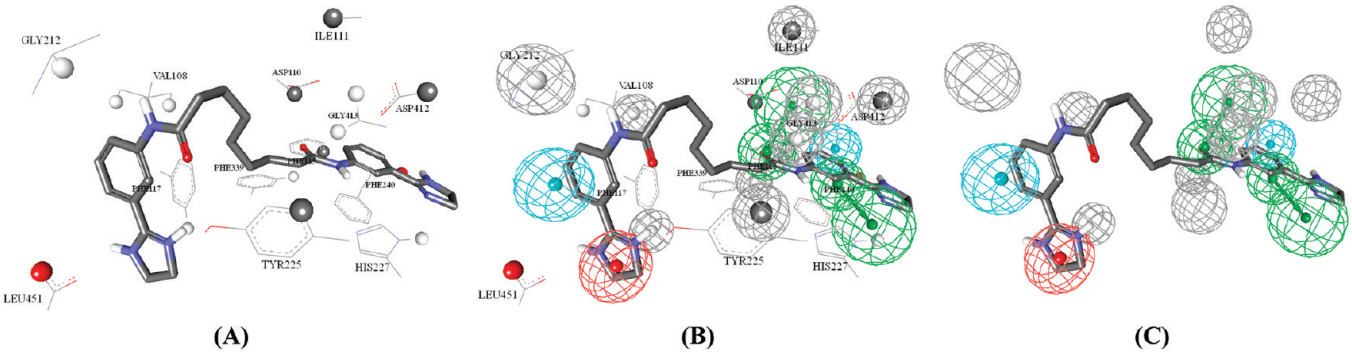


Figure 8. Selective anti-CaNMT hit and how it docks into the binding site via dbCICA model **A-II** docking conditions (Table 1) compared to how it maps **HypoA-II**. (A) Docking **A129** into CaNMT. (B) Docking of **A129** and mapping against **HypoA-II**. (C) Mapping of **A129** against **HypoA-II**.

Table 6. Highest Ranking dbCICA Models for GP Inhibitors, Their Corresponding Parameters, and Statistical Criteria[a]

dbCICA model	ligand ionization state	explicit water[b]	scoring function	contacts distance threshold[c]	number of positive contacts[d]	number of negative contacts[e]	r^2_{35}[f]	r^2_{LOO}[g]	$r^2_{5\text{-fold}}$[h]	F statistic
B−I	ionized	present	Jain	3.5	6	10	0.75	0.72	0.72	106.40
B−II	ionized	present	LIGSCORE2	3.5	5	10	0.73	0.69	0.71	92.65
B−III	ionized	present	PLP1, PMF	3.5	9	10	0.76	0.73	0.73	112.13
B−IV	ionized	absent	PLP2	3.5	9	5	0.76	0.74	0.75	112.42
B−V	unionized	absent	LIGSCORE1	2.5	5	10	0.79	0.76	0.77	130.08

[a] The presented information in this table is extracted from Table G of the Supporting Information except for the F statistic, which is calculated from the correlation connecting −log (IC_{50}) and the contacts sums of the corresponding dbCICA models. [b] Crystalographically explicit water of hydration. [c] Distance thresholds used to define ligand binding site contacts. [d] Optimal number of combined (i.e., summed) bioactivity-enhancing ligand binding site contacts. [e] Optimal number of combined (i.e., summed) bioactivity-disfavoring ligand binding site contacts. [f] Noncross-validated correlation coefficient for 35 training compounds. [g] Cross-validation correlation coefficients determined by the leave-one-out technique. [h] Cross-validation correlation coefficients determined by the leave-20%-out technique repeated 5 times.

docked into the binding site of GP employing LigandFit.[70] Four docking configurations were implemented: ionized vs un-ionized ligands, and hydrous vs anhydrous binding site.

Subsequently, high-ranking docking solutions were scored by six different scoring functions: Jain,[33,37] LigScore1, LigScore2,[70,31] PLP1,[71] PLP2,[41] and PMF.[43−45]

Thereafter, dbCICA modeling was repeated to cover all possible docking combinations, i.e., resulting from the presence (or absence) of crystallographically explicit water molecules within the binding site and the ionization states of the ligands (ionized or un-ionized).

Two distance thresholds, 2.5 and 3.5 Å, were used to determine intermolecular ligand binding site contacts. Accordingly, two corresponding binary contact matrices were generated for each docking solution (see section 4.1.6 of the Experimental Section). A GA-based search was subsequently implemented to search for the best combination (summation) of ligand−receptor intermolecular contacts capable of explaining bioactivity variation across the training compounds. GA was instructed to scan combinations of 2−10 directly proportional intermolecular (positive) contacts followed by 2, 5, or 10 inversely proportional (negative) contacts.

Table 6 shows the docking conditions, distance thresholds, number of positive and negative contacts, and statistical criteria of top-ranking dbCICA models produced for GP inhibitors. The complete list is shown in Table G of the Supporting Information. Apparently, from Table 6 and Table G of the Supporting Information, all different docking settings yielded dbCICA models of relatively good overall qualities with their $r^2_{5\text{-fold}}$ values ≥ 0.58 suggesting the ability of GP inhibitors to assume multiple binding modes within the binding pocket. Still, ligand ionization seems to enhance the qualities of the dbCICA models as four out of the highest-ranking five dbCICA models (Table 6) were based on docking solutions performed on ionized ligands. A similar trend is noticed for explicit water molecules: three out of the top five dbCICA models were based on docking solutions performed on hydrous binding pocket.

The fact that three of the best dbCICA models (i.e., **B−I**, **B−II** and **B−III**, Table 6) were based on the same docking conditions (ionized ligands and hydrous binding site) supports the validity of the corresponding docking conditions and therefore should provide excellent framework for subsequent development into promising pharmacophoric hypotheses for in silico screening.

Table 7 shows amino acid contacts that emerged in the top-deterranking dCICA models. Apparently from the table, some critical amino acid contacts repetitively emerge in high-ranking dbCICA models, which further support their significance. For example, the contact **B**:THR38:CA appears in four out of the five top models, while the amino acids **B**:THR38 and **B**:ARG60 emerged in five and four models, respectively, albeit via different atoms.

Accordingly, it was decided to generate five pharmacophore models from the five highest-ranking dbCICA models. For example, Figure 9 shows how dbCICA model **B−III** was translated into corresponding pharmacophoric model **HypoB-III** (Figure 10). Initially, the binding pocket was annotated by

dx.doi.org/10.1021/ci100368t |J. Chem. Inf. Model. 2011, 51, 647–669

Table 7. Critical Binding Site Contact Atoms Proposed by High Ranking dbCICA Models for GP Inhibitors

	favored contact atoms (positive contacts)[b]		disfavored contact atoms (negative contacts)[c,e]		
dbCICA model[a]	amino acids (or water) and corresponding atom identities[c]	weights[d]	from chain A in GP (PDB code: 1L5R)	from chain B in GP (PDB code: 1L5R)	from explicit water[f]
B−I	**A**:ARG60:NH1	2		GLY186:CA	
	A:LYS191:C	3			
	A:LYS191:HD1	3	LEU39:HA LYS191:CE THR38:C	LYS191:CD LYS191:HD2 LYS191: HE1 TYR185:HD2	H_2O56:O H_2O230:H2
	B:ARG60:CZ	2			
	B:THR38:O	3			
	B:THR38:CA	2			
B−II	**A**:PRO229:CD	2			
	A:PRO188:O	1			
	B:ARG60:NH2	2	LEU39:HA	GLU190:C LYS191:CE LYS191:HD1 LYS191:HD2 TYR185:N	H_2O13:H1 H_2O56:H2 H_2O231:H1 H_2O509:H1
	B:THR38:CA	2			
	H_2O379:H1	3			
B−III	**A**:ARG60:HH12	1			
	A:LEU63:HB2	1			
	A:TRP189:HA	1			
	A:TRP67:CE3	1			
	B:ARG60:HH11	3	HIS57:NE2 LYS191:HZ1 THR38:C	GLY186:CA LYS191:HZ2 PRO188:CA	H_2O4:O H_2O56:H1 H_2O56: H2 H_2O260:H1
	B:GLY186:HA2	3			
	B:THR38:CA	3			
	B:TYR185:HB2	3			
	B:TYR185:O	2			
B−IV	**A**:GLU190:C	2			
	A:TRP189:C	3			
	B:ARG60:CZ	2			
	B:ARG60:HH11	3			
	B:LEU39:HD13	1	ARG60:HE TRP189:CA	ARG60:HH12 LYS191:HZ2 THR38:HG23	
	B:PHE53:HZ	1			
	B:THR38:CA	3			
	B:TYR185:HA	1			
	B:TYR185:HB2	2			
B−V	**A**:LYS191:HB1	2			
	B:HIS57:NE2	2	LYS191:HZ2 PRO194:HG2 SER192:HB1 VAL64:HG23	GLU190:HG1 GLU190:HG2 GLU190: OE2 LEU39:HD12 LYS191:HD2 PRO188:HB2	
	B:PRO188:HD2	1			
	B:THR38:HA	2			
	B:PHE53:HE2	2			

[a] Model numbers are as in Table 6. [b] Bioactivity-enhancing ligand binding site contacts. [c] Binding site amino acids and their significant atomic contacts. Atom codes are as provided in the PDB format (e.g., A:TRP189:CA encodes for peptide chain A, tryptophan number 189, carbon atom A). Hydrogen atoms were labeled as coded by DS2.0. [d] Weight of corresponding contact atom. It points to the times the particular contact appeared in the final dbCICA model (see point (v) in section 4.1.6 of the Experimental Section). [e] Bioactivity-disfavoring ligand binding site contacts. [f] Crystallographically explicit water.

rendering significant contact atoms in spherical forms (see Figure 9a). Subsequently, potent, well-behaved docked compounds, i.e., of least difference between experimental and fitted bioactivities, were placed into the binding pocket to position appropriate pharmacophoric features onto common aligned chemical functionalities among these compounds, as in panel (d) of Figure 9. The pharmacophoric features were placed in such a way to highlight the interactions encoded by the critical contacts (as in the previous case in section 2.4 of the Results and Discussion).

A similar strategy was implemented for the development of four other anti-GP pharmacophore models, namely, **HypoB-I**, **HypoB-II**, **HypoB-IV**, and **HypoB-V** (Figures **A**−**J** of the Supporting Information) corresponding to dbCICA models

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

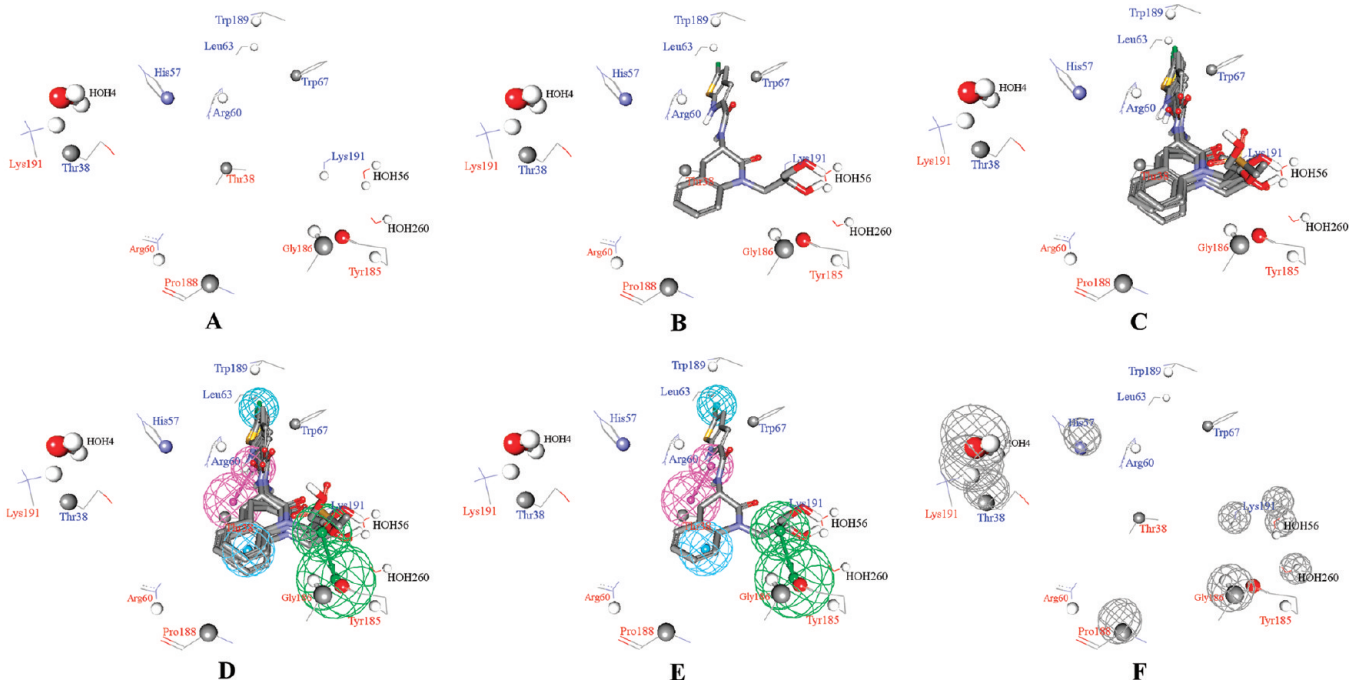


Figure 9. Steps for manual generation of anti-GP binding hypothesis **HypoB-III** as guided by dbCICA model **B−III** (Tables 6 and 7). (A) The binding site moieties in dbCICA model **B−III** with significant contact atoms shown as spheres. Blue-colored and red-colored residues refer to amino acids originating from chains A and B, respectively, of GP protein (PDB code: 1L5R, resolution 2.1 Å). (B) The docked pose of the well-behaved compound **B35** (IC_{50} = 12 nM) within the binding pocket. (C) The docked poses of the well-behaved and potent compounds **B7, B12, B22, B29**, and **B35**. (D) Manually placed pharmacophoric features onto chemical moieties common among docked well-behaved GP inhibitors **B7, B12, B22, B29**, and **B35**. (E) The docked pose of **B35** and how it relates to the proposed pharmacophoric features. (F) Exclusion spheres fitted against binding site atoms showing negative correlations with bioactivity.

B−I, **B−II**, **B−IV**, and **B−V** (Tables 6 and 7), respectively. Table H of the Supporting Information shows the *X*, *Y*, *Z* coordinates of the generated pharmacophores.

2.5.1. Validation of Optimal Anti-GP dbCICA Models. Similar to the CaNMT case, we implemented three validation methods to assess the dbCICA models of GP inhibitors, namely, CoMFA, self-docking, and ROC analyses.

(1). CoMFA Modeling. It was decided to assess whether molecular alignments corresponding to highest ranking dbCICA models yield self-consistent and predictive CoMFA models.[7,59,60] For comparison purposes, a similar assessment was performed for molecular alignments of low-ranking dbCICA models. Table 8 shows the statistical criteria of the resulting CoMFAs.

Clearly from Table 8, the statistical criteria of the first three dbCICA models (i.e., **B−I** to **B−III**, Table 6) correlate nicely with those of their CoMFA counterparts. However, CoMFAs corresponding to dbCICA models **B−IV** and **B−V** were weaker, particularly in their r^2_{PRESS} values. We believe this can be explained because of the limited number of training compounds (28 compounds), which makes it harder on the genetic-partial least-squares-based search (G-PLS) implemented in CoMFA modeling to consistently filter out energy field points of significant correlation with bioactivity, particularly because of the large number of field points generated during CoMFA modeling (thousands). Nevertheless, the low-ranking dbCICA models coincided with low quality CoMFA models (failed r^2_{PRESS} values, bottom of Table 8).

The general agreement between CoMFA and dbCICA in GP inhibitors case provides additional evidence on the capacity of dbCICA modeling as validation technique in docking studies.

(2). Self-Docking. To assess the similarity between ligand poses and conformers produced by the docking conditions of dbCICA models **B−I** to **B−V** (Table 6) and the crystallographic structure of corresponding bound ligands, we compared the docked pose of the ligand CP403700 employing docking conditions of **B−I** to **B−V** with the cocrystallized pose of this ligand (PDB code: 1L5R, resolution 2.1 Å). Table 9 shows the rms difference between the experimental structure and each of the docked structures. Clearly from the table the top three docking conditions (**B-1** to **B−III**) succeeded docking the ligand in close conformers and poses to the cocrystallized pose with rms difference of 0.5 Å. However, docking conditions corresponding to the dbCICA models **B−IV** and **B−V** deviated significantly from the cocrystallized pose probably because both docking settings are based on anhydrous binding pocket in contrast to the hydrous binding site in the crystallographic structure. Still, agreement between dbCICA-based docking conditions and crystallographic structures under equivalent conditions supports the validity of dbCICA modeling.

(3). Receiver Operating Characteristic Curve Analyses of HypoB-I to HypoB-V. Table 10 and Figure 11 (and Figures K and L of the Supporting Information) show ROC performances of **HypoB-I** to **HypoB-V** and their sterically refined versions (see sections SM-1 and SM-2 and Table 1 of the Supporting Information for ROC analysis and steric refinement of pharmacophores).[62,78−81,112−115] Clearly from Table 10 and related figures, all dbCICA-based pharmacophores shared good overall ROC performances, which adds further support to their corresponding dbCICA models. Moreover, steric refinement of the pharmacophores significantly improved their ROC

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

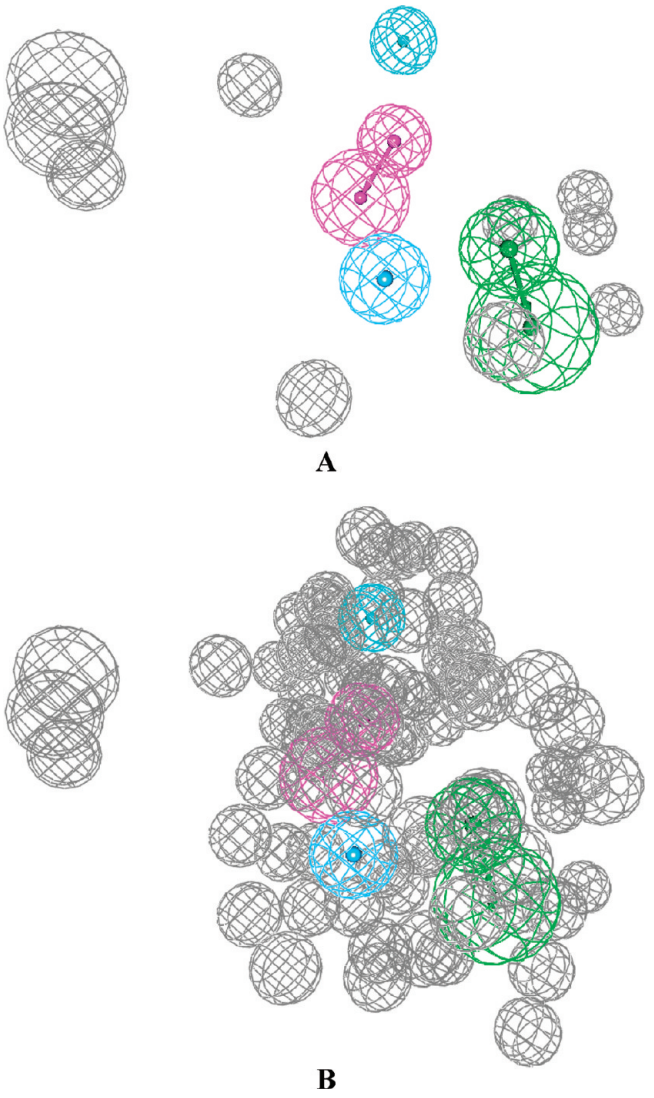


Figure 10. (A) **HypoB-III** pharmacophoric features and exclusion spheres. Light blue spheres represent hydrophobic features, dark blue spheres represent hydrophobic aromatic features, vectored green spheres represent hydrogen bond acceptor features, vectored pink spheres represent hydrogen bond donors, and gray spheres represent exclusion regions. (B) Refined **HypoB-III** with 68 added exclusion spheres.

performances. However, the refined version of **HypoB-III** (Figure 10) illustrated particularly excellent ROC qualities (Figure 11) prompting us to use it as a 3-D search query in subsequent in silico screening for anti-GP leads.

2.5.2. In Silico Mining for New GP Inhibitors. In silico screening was conducted employing **HypoB-III** as a 3-D search query against the National Cancer Institute list of compounds (NCI, includes 238,819 compounds). NCI hits were subsequently filtered by a molecular weight threshold of 550 Da in order to remove large, nondrug-like compounds. Another filter was performed based on the number of rotable bonds. The resulting hits were docked into GP protein (PDB code: 1L5R) employing the same docking conditions of dbCICA model **B−III** (i.e., ionized ligands with protein hydration water and employing PLP1 scoring function, Tables 6 and 7). The resulting docked poses were subsequently analyzed for critical contacts according to dbCICA model **B−III** (Table 7), and the sums of critical contacts for each hit compound were used to rank the corresponding hits and prioritize subsequent in vitro testing. Table 11 and Figure 12 show the highest ranking hits and their experimental in vitro bioactivities. Figure 13 shows how an anti-GP hit docks into the binding pocket of GP and how the docked conformer maps **HypoB-III**.

3. CONCLUSIONS

In the current project, we introduced a novel methodology to validate and identify optimal docking configuration(s), namely, docking-based comparative intermolecular contacts analysis (dbCICA). Two cases studies were presented: candida NMT and GP. Our novel approach proved to be useful to validate docking studies and as a means to screen virtual compound libraries for new hits.

4. EXPERIMENTAL SECTION

4.1. Molecular Modeling. *4.1.1. Software and Hardware.* The following software packages were utilized in the present research.

- CS ChemDraw Ultra (Version 11.0), Cambridge Soft Corp., USA.
- MarvinView (Version 5.1.4), ChemAxon Ltd., USA.
- DiscoveryStudio (DS 2.5), Accelrys, Inc., USA.
- Ligandfit within CERIUS2 (Version 4.10), Accelrys, Inc., USA.
- Libdock within Discovery Studio (Version 2.0), Accelrys, Inc., USA.
- FRED (Version 2.1), OpenEye Scientific Software, Inc., USA.
- OMEGA (Version 2.1.0), OpenEye Scientific Software, Inc., USA.
- MATLAB (Version R2007a), The MathWorks, Inc., USA.
- Catalyst (version 5.11), Accelrys, Inc., USA.

4.1.2. Data Set

4.1.2.1. Case A: CaNMT. The structures of 108 CaNMT inhibitors were collected from the literature.[64−68] They were carefully collected such that they were all bioassayed employing similar conditions in order to allow proper structure−activity correlation. The in vitro bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of CaNMT enzyme by 50%, i.e., IC_{50}. Table A of the Supporting Information shows the structures and IC_{50} values of the collected inhibitors. The logarithm of measured IC_{50} (μM) values were used in structure−activity correlation, thus correlating the data linear to the free energy change. In cases where IC_{50} values were expressed as being higher than certain values, e.g., **A10**, **A14**, **A15**, **A16**, **A17**, **A18**, **A19**, **A20**, **A26**, **A28**, **A29**, **A31**, **A32**, **A33**, **A41**, **A42**, **A65**, and **A71** (having $IC_{50} > 10$ μM), we assumed the exact value (i.e., 10 μM). These assumptions are necessary to allow statistical correlation. The logarithmic transformation of IC_{50} values should minimize any potential error resulting from this assumption as well as from minor discrepancies in bioassay conditions among different authors.

4.1.2.2. Case B: GP. The structures of 45 GP inhibitors (**B1**−**B45**, Table F of the Supporting Information) were assembled from published literature.[116,117] The in vitro bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of GP enzyme by 50%,

dx.doi.org/10.1021/ci100368t |J. Chem. Inf. Model. 2011, 51, 647–669

Table 8. Statistical Results of the CoMFA Models Obtained via Anti-GP dbCICA-Based Docking−scoring Combinations (as in Table 6)

			statistical criteria						
dbCICA models[a]		docking conditions[a]	scoring function[a]	terms[b]	PC[c]	r^{2d}	r^2_{LOO}[e]	r^2_{BS}[f]	r^2_{PRESS}[g]
highest[h] ranking	**B−I**[h]	ionized ligands: hydrous binding pocket	Jain	7	5	0.91	0.72	0.89	0.56
	B−II	ionized ligands: hydrous binding pocket	Ligscore2	7	5	0.90	0.56	0.76	0.68
	B−III	ionized ligands: hydrous binding pocket	PLP1 or PMF	7	5	0.89	0.74	0.87	0.50
	B−IV	ionized ligands: anhydrous binding pocket	PLP2	6	3	0.78	0.49	0.75	−1.05 (0.56)[j]
	B−V	un-ionized ligands: anhydrous binding pocket	Ligscore1	4	3	0.79	0.59	0.62	−0.14
low ranking[i]		un-ionized ligands: hydrous binding pocket	PMF	5	3	0.80	0.32	0.61	−0.22
		un-ionized ligands: hydrous binding pocket	PLP2	6	4	0.88	0.41	0.64	−0.26
		un-ionized ligands: anhydrous binding pocket	Ligscore2	7	5	0.85	0.24	0.53	0.10

[a] dbCICA models and corresponding docking−scoring conditions were selected from Tables 6 and 7. [b] Number of CoMFA descriptors in the best 3-D QSAR model. [c] Number of principal components (latent variables) in the best 3-D QSAR model. [d] Noncross-validated correlation coefficient for 28 training compounds. [e] Cross-validation correlation coefficients determined by the leave-one out technique. [f] Bootstrapping correlation coefficient. [g] Predictive r^2 determined for seven test compounds. [h] Docking−scoring conditions of highest ranking dbCICA models. Models numbered as in Tables 6 and 7. [i] Docking−scoring conditions of low-ranking dbCICA models selected from Table G of the Supporting Information. [j] r^2_{PRESS} value after removing one outlier (**B15**, Table F of the Supporting Information).

Table 9. Root Mean Square (rms) Differences between Co-Crystallized Pose of Ligand CP403700 (PDB code: 1L5R) and Its Docked Comformers and Poses Based on Docking Conditions of High-Ranking dbCICA Models

dbCICA models[a]	docking conditions[a]	scoring function[a]	rms (Å)
B−I	ionized ligands: hydrous binding pocket	Jain	0.50
B−II	ionized ligands: hydrous binding pocket	Ligscore2	0.50
B−III	ionized ligands: hydrous binding pocket	PLP1	0.50
B−IV	ionized ligands: nhydrous binding pocket	PLP2	8.0
B−V	unionized ligands: anhydrous binding pocket	Ligscore1	3.0

[a] dbCICA models and corresponding docking−scoring conditions were selected as in Tables 6 and 7.

Table 10. Receiver Operating Characteristic (ROC) curve Performances of Pharmacophore Models generated from high-Ranking dbCICA Models Generated for GP Inhibitors

pharmacophore model	AUC[a]	ACC[b]	SPC[c]	TPR[d]	FNR[e]
HypoB-I	76	96	97	61	3
Refined-HypoB-I	97	96	99	14	1
HypoB-II	67	96	97	64	3
Refined-HypoB-II	97	96	99	29	1
HypoB-III	76	96	98	46	2
Refined-HypoB-III	98	96	99	29	1
HypoB-IV	91	96	99	29	1
Refined-HypoB-IV	89	96	100	7	0
HypoB-V	77	96	100	11	0
Refined-HypoB-V	85	96	98	43	2

[a] AUC: area under the curve (%). [b] ACC: overall accuracy (%). [c] SPC: overall specifity (%). [d] TPR: overall true positive rate (%). [e] FNR: overall false negative rate (%).

i.e., IC_{50}. Table F of the Supporting Information shows the structures and IC_{50} values of the collected inhibitors. Only 35 compounds were docked and considered in dbCICA modeling (**B1**−**B35**, Table F of the Supporting Information) because they were bioassayed by identical procedure. The remaining compounds (**B36**−**B45**, Table F of the Supporting Information) were used for steric refinement of generated pharmacophore models (see section SM-1 and Table 1 of the Supporting Information).

Racemic inhibitors were assigned arbitrary absolute chiral configurations for subsequent modeling. This assumption is based on the fact that the enantiomeric pairs of most ligands behaved similarly under different docking−scoring approaches, i.e., they yielded visually similar docked conformers and poses under a particular docking condition. However, to avoid docking discrepancies in cases of diasteriomeric molecules, training compounds of more than one asymmetric center were docked into GP in their absolute configurations, i.e., **B31**, **B32**, **B33**, **B34**, and **B35** (Table F of the Supporting Information).

4.1.3. Preparation of Ligands. The two-dimensional structures of collected inhibitors (**A1**−**A108** in Table A and **B1**−**B35** in Table F of the Supporting Information) were sketched in ChemDraw Ultra (Version 11.0). Two protonation states were assumed for each inhibitor, ionized and un-ionized, as guided by MarvinView. In the ionized forms, amino substituents ($pK_a \approx 9.0-9.5$) were protonated and given formal positive charges at the appropriate atoms, while carboxylic acids were deprotonated and given formal negative charges at the appropriate atoms. The structures were subsequently converted into reasonable three-dimensional representations employing the rule-based conformational methods implemented in DS2.5 and were saved in SD format for subsequent docking experiments.

dx.doi.org/10.1021/ci100368t |J. Chem. Inf. Model. 2011, 51, 647–669

4.1.4. Preparation of Crystal Structures. The 3-D coordinates of both CaNMT and GP were retrieved from the Protein Data Bank (CaNMT PDB code: 1IYK, resolution: 2.3 Å, GP PDB code: 1L5R, resolution: 2.1 Å). Hydrogen atoms were added to the protein utilizing DS 2.0 templates for protein residues. Gasteiger-Marsili charges were assigned to the protein atoms as implemented within DS 2.5.[69]

The protein structures were utilized in subsequent docking experiments without energy minimization. Explicit water molecules were either kept or removed according to the required docking conditions, i.e., docking in the presence or absence of explicit water molecules.

4.1.5. Docking Configurations

(1). LigandFit Docking and Scoring. LigandFit considers the flexibility of the ligand and treats the receptor as rigid. There are two steps implemented in the LigandFit process.

(A) Defining the location(s) of potential binding site(s).[70,71] In the current docking experiments, the binding site was generated from the cocrystallized ligand (MIM) within the targeted protein. The cocrystallized myristoyl-CoA was considered part of the protein and was left unaltered during docking.

(B) Docking the ligands in the binding site.[70,71] In LigandFit, docking is composed of few substeps: (i) Conformational search of flexible ligands employing Monte Carlo randomized process. (ii) Pose and conformation selection based on shape similarity with the binding site. (iii) Candidate conformers and poses exhibiting low shape discrepancy are further enrolled in calculation of the dock energies. (iv) Each docked conformation and pose is further fitted into the binding pocket through a number of rigid-body minimization iterations. (v) Docked conformers and poses that have docking energies below certain user-defined threshold are subsequently clustered according to their rms similarities. Representative conformers and poses are then selected, further energy-minimized within the binding site, and saved for subsequent scoring.

In the current docking experiments, all 108 ligands in their ionized and un-ionized forms were docked into the binding site

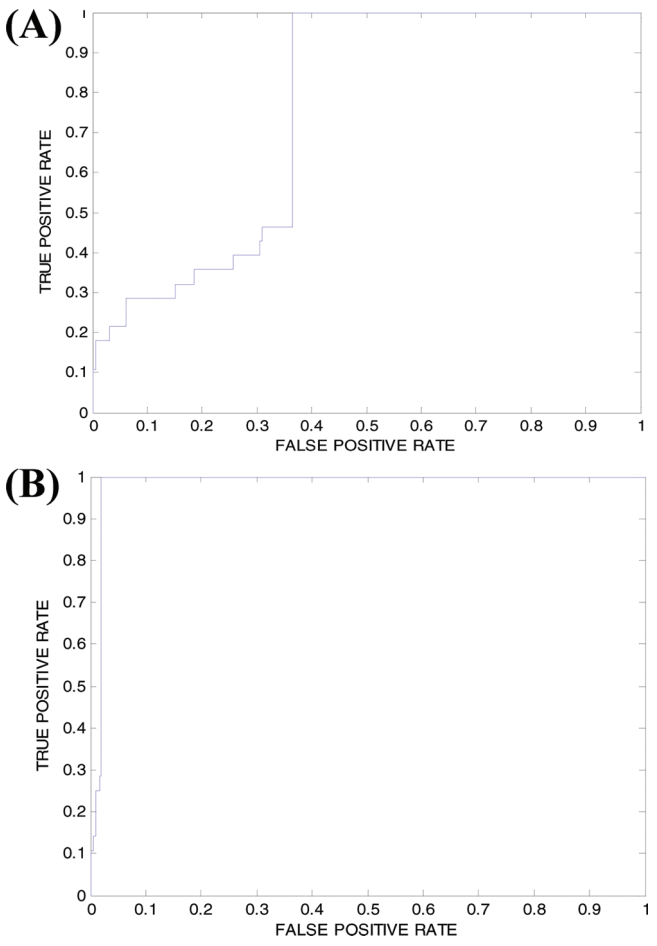


Figure 11. Receiver operating characteristic (ROC) curves of (A) dbCICA-based GP pharmacophore **HypoB-III** and (B) sterically **Refined-HypoB-III**.

Table 11. High-Ranking Hits Captured by Refined-HypoB-III, Their Corresponding Summation Contacts According to dbCICA model B−III, and in Vitro Bioactivities

hit[a]	NCI code	dbCICA contacts[b]	% inhibition of GP at 10 μM concentration[c]
B48	34304	14	17
B49	69411	14	16
B50	71881	14	21
B51	80728	17	12
B52	88631	22	10
B53	89532	18	11
B54	110617	17	20
B55	119608	16	20
B56	127121	15	15
B57	127122	15	21
B58	204614	18	33 (IC_{50} = 500 μM)[d]
B59	351585	16	27
B60	369110	17	10
B61	646817	13	10
B62	648593	13	10

[a] Hit numbers as in Figure 12. [b] Contact summations according to dbCICA model **B−III** (Tables 6 and 7). [c] Inhibition percentage values were measured as duplicated readings. [d] Regression coefficient of the dose/inhibition plot is 0.99.

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

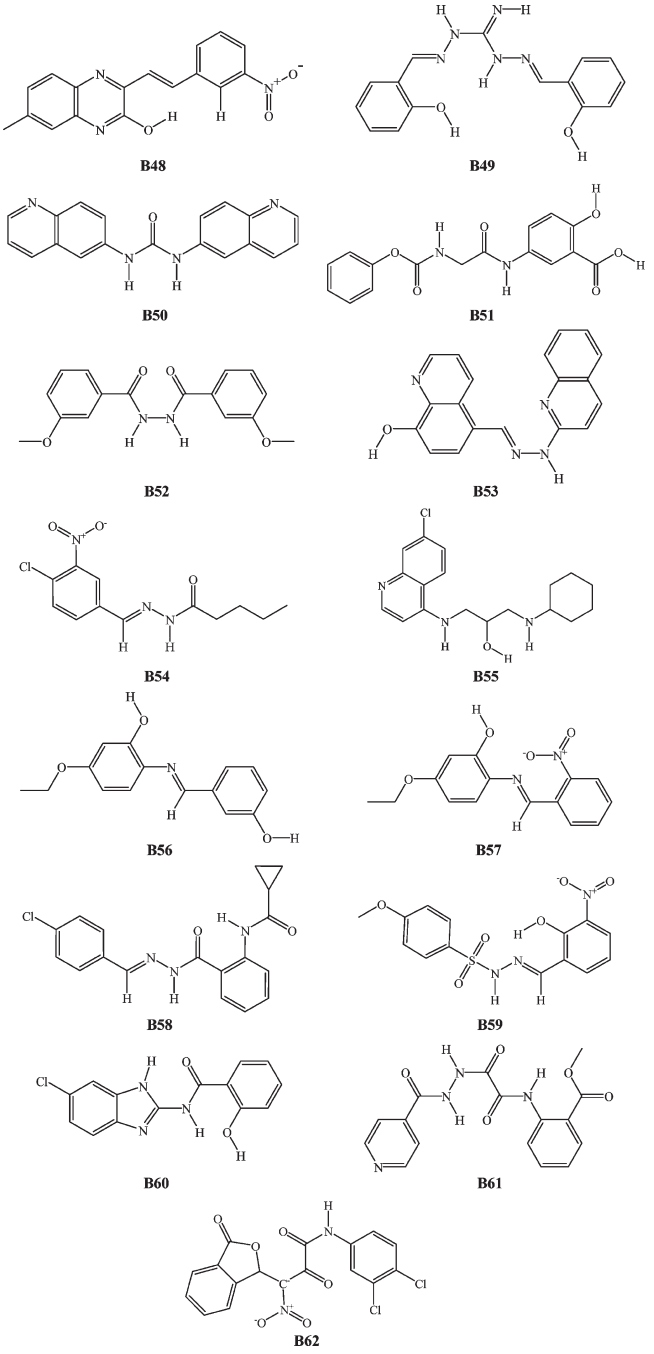


Figure 12. Chemical structures of the tested highest ranking hits captured by pharmacophoric **HypoB-III**.

in the presence and absence of explicit water molecules, employing the following docking configurations.

- Monte Carlo search parameters: number of trials = 50000 and search step for torsions with polar hydrogens = 30.0°.
- The rms threshold for ligand-to-binding site shape match was set to 2.0 Å, employing a maximum of 3.0 binding site partitions
- Interaction energy parameters: The interaction energies were assessed employing CFF force field (version 1.02) with a nonbonded cutoff distance of 10.0 Å and distance dependent dielectric. An energy grid extending 5.0 Å from the binding site was implemented. The interaction energy was estimated by a trilinear interpolation value using soft potential energy approximations.[70]
- Rigid body ligand minimization parameters: 30 steepest descend followed by 60 BFGS minimization iterations were applied to every orientation of the docked ligand. The best 10 poses were further energy minimized within the binding site for a maximum of 300 rigid body iterations.
- High-ranking docked conformers and poses were scored using seven scoring functions: Jain,[37] LigScore1, LigScore2,[70,31] PLP1,[71] PLP2,[41] PMF, and PMF04.[44,45,47] Considering each scoring function in turn, the highest scoring docked conformer and pose was selected for each inhibitor for subsequent 3-D QSAR modeling. This resulted in seven sets of 108 docked molecules with scores corresponding to each scoring function. However, the docking and scoring cycle was repeated 4 times (2 × 2) to cover the different combinations of docking conditions, i.e., explicit water molecules (hydrous and anhydrous binding site), and ligand ionization (ionized and un-ionized).

LigScore1 and LigScore2 scores were calculated employing CFF force field (version 1.02) and using grid-based energies with a grid extension of 7.5 Å across the binding site. PMF scores were calculated employing cutoff distances of 12.0 Å for carbon−carbon interactions and other atomic interactions, while PMF04 scores were calculated employing cutoff values of 6.0 and 9.0 Å for carbon−carbon interactions and other atomic interactions, respectively.

(2). FRED Docking and Scoring. FRED (fast rigid exhaustive docking) employs a docking strategy that exhaustively scores all possible positions and orientations of each ligand in the active site.[26] The ligand conformers and protein structure are treated as rigid during FRED docking process. The exhaustive search is based on rigid rotations and translations of each conformer within the binding site defined by a box created around a co-crystallized ligand within the binding site. FRED filters poses ensemble by rejecting the ones that clash with the protein or those that do not have enough contacts with the receptor. The final poses can then be scored or rescored using one or more scoring functions.

The conformational spaces of the docked compounds were explored using OMEGA (optimized ensemble generation application) program.[26] The algorithm implemented in OMEGA dissects the molecules into fragments and uses fragment templates to build a seed conformation. Next, OMEGA begins torsion search with an assessment of freely rotable bonds. We implemented default parameters for generating different conformers employing OMEGA.

The following FRED docking parameters were employed.

- Binding Box: Defined employing the c-ocrystallized ligand (MIM) in the CaNMT crystal structure (1IYK) from the protein data bank.
- Addbox: This optional parameter adjusts the geometry of the box defining the active site by extending each edge of the box by the specified number of angstroms. In this experiment, this value was set to 3.00 (default = 0).
- Num_poses: This parameter specifies the number of high-ranking poses to be returned by the exhaustive search. These poses will be scored by the selected scoring functions.

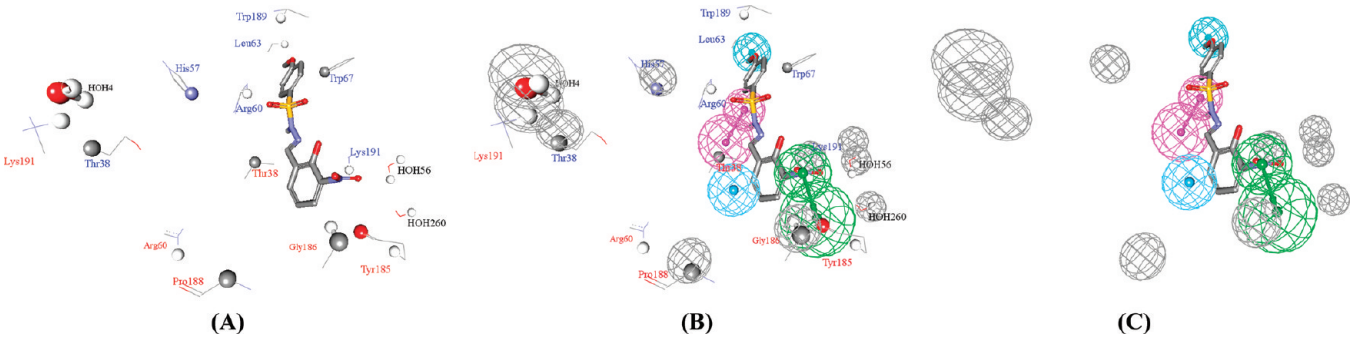


Figure 13. Selective anti-GP hit and how it docks into the binding site via dbCICA model **B**−**III** docking conditions (Table 6) compared to how it maps **HypoB-III**. (A) Represents docking **B59** into GP. (**B**) Docking of **B59** and mapping against **HypoB-III**. (**C**) Mapping of **B59** against **HypoB-III**.

In this experiment, this value was set to 1000 (default = 100).

- Num_alt_poses: This flag specifies how many high-ranking alternate poses, in addition to the top scored pose, will be delivered to the user. Values are between 0 and 99. In the current experiment this value was set to be 10.
- The docked poses were scored by seven scoring functions, namely, Shapegauss,[72] Chemgauss2,[26] PLP,[73] Consensus,[74] Chemscore,[75] Screenscore,[76] and Zapbind.[77] The highest-ranking poses were retained for subsequent 3-D QSAR. This resulted in seven sets of 108 docked molecules with scores corresponding to each scoring function. However, the docking and scoring cycle was repeated 4 times (2 × 2) to cover the different combinations of docking conditions, i.e., explicit water molecules (hydrous and anhydrous binding site), and ligand ionization (ionized and un-ionized).

(3). LibDock Docking. LibDock docks ligands (after removing hydrogen atoms) into a binding site guided by binding hotspots. It aligns docked ligand conformations to polar and apolar receptor interactions sites, i.e., hotspots. Conformations can be either precalculated or generated on the fly. Because some of the output poses may have hydrogen atoms in close proximity to the receptor, a CHARMm minimization step can be optionally enabled to further optimize the docked poses.[27,28]

LibDock performs the following steps using a set of pregenerated ligand conformations and a receptor with a specified binding site:[27,28] (i) Remove hydrogen atoms. (ii) Rank ligand conformations and prune by solvent accessible surface area (SASA). (iii) Find hotspots using a grid placed into the binding site and using polar and apolar probes. The numbers of hotspots are pruned by clustering to a user defined value. (iv) Dock ligand poses by aligning to the hotspots. This is performed by using triplets (i.e., three ligand atoms are aligned to three receptor hotspots). Poses which result in protein clashes are removed. (v) A final BFGS pose optimization stage is performed using a simple pairwise score. The top scoring ligand poses are retained. (vi) Hydrogen atoms are then added back to the docked ligands. (vii) Optionally, CHARMm minimization can be carried out to reduce steric clashes caused by added hydrogen atoms.

In the current docking experiments we employed the following parameters.

- Binding site sphere radius of 9 Å surrounding the center of the cocrystallized ligand (MIM).
- Maximum number of receptor hotspots = 100.
- Ligand-to-hotspots matching rmsd tolerance value was set to 0.25 Å.
- Maximum number of saved poses for each ligand = 100.
- Maximum number of poses saved for each ligand during hotspots matching before entering the final pose minimization = 100.
- Minimum LibDock score (poses below this score are not reported) = 100.
- Fraction of reported top scoring poses = 0.5.
- Maximum number of rigid body minimization steps during final pose optimization phase (using BFGS method) = 50.
- Maximum number of evaluated poses for each conformation = 30.
- Maximum number of steric clashes allowed before the pose−hotspot alignment is terminated (specified as a fraction of the heavy atom count) = 0.10.
- Cluster similarity cutoff value = 0.5 Å (docked poses are rigid-body minimized and clustered using this cutoff value).
- Maximum value for nonpolar solvent accessible surface area for a particular pose to be reported as successful = 15.0 $Å^2$.
- Maximum value for polar solvent accessible solvent area for a particular pose to be reported as successful = 5.0 $Å^2$.
- Number of grid points used for calculating solvent accessible surface area = 18.
- Conformation generation method: The CATALYST module CATCONFIRM implemented in DS 2.0 was implemented employing the BEST conformation generation option to ensure the best coverage of the compound's conformational space. Maximum number of conformations to be generated per ligand = 255 not exceeding an energy threshold of 20 kcal/mol from the most stable conformer. No final ligand minimization was implemented (i.e., in the binding pocket).
- The docked poses were scored employing the same seven scoring functions that were implemented in LigandFit docking experiment and employing identical parameters.

4.1.6. Docking-Based Comparative Molecular Contacts Analysis (dbCICA). The methodology of dbCICA can be described in the following sequential steps.

(i) High-ranking docked pose and conformer of each ligand, based on a particular docking condition, is evaluated to identify its closest atomic neighbors in the binding pocket. Intermolecular atomic neighbors closer than

dx.doi.org/10.1021/ci100368t |*J. Chem. Inf. Model.* 2011, 51, 647–669

(or equal to) certain predefined distance threshold are assigned an intermolecular contact value of "one", otherwise they are assigned a contact value of "zero". For example, if atom A in the docked ligand is positioned close to atom B in the binding pocket at a distance shorter that the predefined threshold, then this contact is assigned a value of 1. Distance assessment is performed automatically employing the intermolecular monitor implemented in Discovery Studio Version 2.0.[61] Eventually, this step yields a two-dimensional matrix with row labels corresponding to docked ligands (i.e., according to the particular docking−scoring configuration) and column labels corresponding to different binding site atoms. The matrix is filled with binary code, whereby "zeros" correspond to interatomic distances above the predefined threshold and "ones" for distances below (or equal) the predefined threshold. In the CaNMT study, three distance thresholds were implemented, 3.5, 2.5, or 2.0 Å, while in GP study two distance thresholds were implemented, 3.5 or 2.5 Å. Therefore, three (CaNMT) or two (GP) binary matrices (corresponding to each distance threshold) are constructed for each docking configuration (i.e., combination of docking engine, scoring function, ligand ionization state, and binding site hydration status).

(ii) Each individual column in the matrix is regressed against the corresponding molecular bioactivities, i.e.,−log (IC_{50}). Columns that exhibit negative correlation with bioactivity are inverted, i.e., zeros are converted to ones and vice versa, and excluded from the subsequent step.
After excluding inverted columns (negative contacts or exclusion volumes), the resulting binary matrix (which is composed from positively correlated contact columns with bioactivity) is then subjected to genetic algorithm (GA)-based search for optimal summation of contacts columns capable of explaining bioactivity variation: In this step, GA relies on the evolutionary operations of "crossover and mutation" to select optimal combination of columns that have their summation values collinear with bioactivity variation across training compounds (see GA parameters below). The best column summation model (single model) is selected as representative dbCICA model.
GA can be instructed to output best dbCICA models resulting from any predefined number of ligand−receptor intermolecular positive contacts, e.g., best dbCICA models resulting from sets of 2 or 3 or 4 or 5, etc... concomitant contacts (i.e., summed contact columns). In both current projects, we instructed GA to search for the best dbCICA models resulting from two contacts and repeat the scan to identify the best summation models for 3, 4, 5, 6, 7, 8, 9, and 10 contacts. Each set of summed contacts is treated independently to identify the corresponding dbCICA model in each case.

(iii) The dbCICA algorithm has the option of allowing any particular positive contact column to emerge up to three times in the optimal summation model, i.e., it allows variable weights for contacts. This is performed by implementing dual-valued genes in the GA, in which every gene encodes for both the corresponding contacts column number and its weight. Column weights are initially randomly distributed in the first generation and subsequently subjected to mutation only (not crossover) in GA. This option was allowed in the current project to identify intermolecular contacts of higher weights or contributions in the optimal dbCICA models.

(iv) After identifying optimal summation model(s) based on positive contacts (proportional to bioactivity), dbCICA implements GA to search for optimal summation model resulting from combining inverted columns [negatively proportional to bioactivity, see step (ii)] with the optimal positive summation model(s). The user has the option of choosing any number of negative contacts (excluded volumes or steric clashes) to emerge in the final dbCICA model. In the current experiment, we implemented two exclusion settings: either five or ten negative contacts were allowed.

4.1.7. Genetic Algorithm Implementation in dbCICA Modeling. The GA toolbox within MATLAB (Version R2007a) was adapted by implementing the following four basic components: creation function, crossover function, mutation function, and fitness function.

The creation function randomly generates a population of chromosomes of a predefined size (number of summed contacts columns, as mentioned in step (ii) in section 4.1.6 of the Experimental Section) in which every chromosome encodes for certain possible column summation model. Chromosomes differ from one another by the set of summed columns and their weights.

Crossover children are the offspring created by selecting vector entries (i.e., genes) from a pair of individual chromosomes in the first generation and combining them to form two complementary children. While mutation children are those created via applying random changes to corresponding parents, i.e., each single parent chromosome is mutated to give a single child by randomly replacing selected gene in the parent chromosome with another from the chromosome population.

Each chromosome is associated with a fitness value that reflects how good the summation of its encoded genes compares to other chromosomes. The fitness functions in dbCICA can be the correlation coefficient (r^2), leave-one-out r^2, or K-fold r^2.

In the current experiments, we implemented a 5-fold r^2 as fitness criterion. In this procedure, each chromosome is ranked as follows. The training set is divided into two subsets: fit and test subsets. The test subset is randomly selected to represent about 20% of the training compounds. This procedure is repeated over five cycles; accordingly, five test subsets with their complementary fit subsets are selected for each chromosome (i.e., column summation model). The five test subsets should cover about 100% of the training compounds by avoiding selecting the same compound in more than one test subset. The fit sets are then utilized to generate five submodels employing the same chromosome. The resulting submodels are then utilized to predict the bioactivities of the corresponding testing subsets. Finally, the predicted values of all five test subsets are correlated with their experimental counterparts to determine the corresponding 5-fold r^2.

The following parameters were chosen for GA genetic manipulation in dbCICA of CaNMT and GP experiments: size of chromosome population = 200, rate of mating (crossover fraction) = 80%, elite count = 1, maximum number of generations that is needed to exit from GA iteration cycles, and completion of the algorithm = 2000.

On the basis of these settings, the numbers of each type of children in the offspring generation is as follows. There is one elite child (corresponding to the individual in the parents' generation with the best fitness value), and there are 199 individual children other than the elite child. The algorithm rounds 0.8 (crossover fraction) × 199 = 159.2 to 159 to get the number of crossover children and the remaining 40 (i.e., 199−159) are the mutation population. The elite child is passed to the offspring population without alteration.

4.1.8. Molecular Field Analysis. To further validate dbCICA models, we used comparative molecular field analysis (CoMFA) to assess the ability of corresponding docking-based molecular alignments to yield self-consistent 3-D QSAR models.[7,59,60] We used molecular field analysis (MFA) and G/PLS modules implemented in CERIUS2 to perform 3-D QSAR analysis.[7,59,60] The alignments of different inhibitors came directly according to dbCICA-based highest-ranking docking−scoring combinations (Tables 1 and 6). For each alignment, the interaction fields between the ligands and proton (positively charged), hydrogen bond donor/acceptor, and methyl (neutral) probes were calculated employing a regularly spaced rectangular grid of 2.0 Å spacing. The spatial limits of the molecular field were defined automatically and were extended past the van der Waals volume of all the molecules in the *X*, *Y*, and *Z* directions. The ligands were assigned partial charges using the Gasteiger method implemented within CERIUS2. The energy fields were calculated employing the default UNIVERSAL force field (version 1.02) implemented within CERIUS2 and were truncated to ±50 kcal/mol. For the CaNMT experiment, the calculation gave around 4212 variables for each compound (1404 variable/probe), while for the GP experiment the calculation gave nearly 2970 variables for each compound (990 variable/probe).

To derive the best possible 3-D QSAR statistical model for each docking−scoring combination, we used genetic partial least squares (G/PLS) analysis to search for optimal regression equations capable of correlating the variations in biological activities of the training compounds with variations in the corresponding interaction fields.[7,59,60] G/PLS is derived from two methods: genetic function approximation (GFA) and partial least-squares (PLS). GFA techniques rely on the evolutionary operations of "crossover and mutation" to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations (i.e., chromosomes population). Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's "lack-of-fit" (LOF).

G/PLS algorithm uses GFA to select appropriate basis functions to be used in a model of the data and PLS regression as the fitting technique to weigh the basis functions' relative contributions in the final model. Application of G/PLS allows the construction of larger QSAR equations while avoiding overfitting and eliminating most variables.[7,59,60]

Our preliminary diagnostic trials suggested the following optimal G/PLS parameters. Explore linear equations of 4 to 8 terms in the GP case and 10 to 17 in the CaNMT case, at mating and mutation probabilities of 50%, population size = 500, number of generations (iterations) = 30000, and LOF smoothness parameter = 1.0. However, the optimal number of PLS latent variables (or principle components) was determined for each CoMFA model through assessing the corresponding predictive r^2 (r^2_{PRESS}) calculated from an external test sets of 21 CaNMT inhibitors (**A3**, **A6**, **A11**, **A16**, **A20**, **A25**, **A34**, **A41**, **A46**, **A51**, **A58**, **A69**, **A74**, **A75**, **A81**, **A84**, **A87**, **A93**, **A100**, **A101**, and **A108**) or seven GP inhibitors (**B6**, **B10**, **B15**, **B19**, **B26**, **B28**, and **B29** in Table F of the Supporting Information). These molecules were aligned according to the particular docking−scoring configuration, and their activities were predicted by corresponding G/PLS models generated from the corresponding training set and employing a range of three to six latent variables. The optimum number of principle components was defined as the one leading to the highest predictive r^2_{PRESS} and lowest sum of squared deviations between predicted and actual activity values for every molecule in the test set (PRESS). Predictive r^2_{PRESS} is defined as

$$r^2_{\mathrm{PRESS}} = (\mathrm{SD} - \mathrm{PRESS})/\mathrm{SD} \quad (1)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, and PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.[7] All 3-D QSAR models were cross-validated employing leave-one-out (LOO) cross-validation and bootstrapping.

4.1.9. Generation of Pharmacophores Corresponding to Successful dbCICA Models. In order to utilize dbCICA modeling for effective drug discovery, optimal dbCICA models were used to guide development of pharmacophoric models to be subsequently used as search queries for the discovery of new CaNMT or GP inhibitors. Pharmacophoric models were developed through the following steps

(1) Docking configurations that yielded the best dbCICA models were selected. The corresponding docked poses and conformers of the most potent compounds (e.g., IC_{50} < 0.05 μM in CaNMT inhibitors) were retained in the binding pocket while other less potent compounds were discarded.

(2) Subsequently, the best dbCICA model(s) (e.g., models **A-I** or **A-II**, Tables 1 and 2) was used to predict the bioactivity of potent compounds in the binding pocket, i.e., by substituting the number of contacts of each docked compound in the regression equation corresponding to the dbCICA model. Well-behaved potent compounds were retained in the binding pocket for subsequent manipulation. Well-behaved compounds are defined as those training compounds that have their bioactivities well-predicted by the selected optimal dbCICA model, i.e., of least residual difference between fitted and experimental bioactivities as predicted by the particular dbCICA model.

(3) Significant positive contacts in the binding pocket, i.e., those that have weights of 2 or 3, were marked and carefully assessed to identify their closest ligand moieties. Consensus among potent, well-behaved training compounds to place moieties of common physicochemical properties adjacent to significant contact atom (as defined by the dbCICA model) warrants placing a corresponding pharmacophoric feature onto that region. For example, if potent, well-behaved docked compounds agreed on placing aromatic rings adjacent to certain dbCICA significant contact point (within the predefined distance threshold, see point (i) in section 4.1.6 of the Experimental Section), then a hydrophobic aromatic feature is placed on top of

the aromatic rings. The pharmacophoric query features were added manually from the DS 2.0 feature library and employing default feature radii (1.6 Å).

(4) Finally, to account for the steric constraints of the binding pocket, binding site atoms that exhibit contacts of negative correlations with bioactivity, i.e., those inverted in step (ii) in section 4.1.6 of the Experimental Section, were marked and used as centers for exclusion spheres. Negative contacts identify spaces occupied by docked conformers and poses of inactive compounds and free from active ones and therefore can be filled with exclusion volumes. Exclusion spheres were added manually from DS 2.0 feature library and employing default feature radii (1.2 Å).

4.1.10. In Silico Screening of the NCI Database for New CaNMT and GP Inhibitors. CaNMT **HypoA-II** and GP **Refined-HypoB-III** were employed as 3-D search queries to screen the national Cancer Institute (NCI) list of compounds (238,819 compounds). The screening was performed employing the "Best Flexible Database Search" option implemented within CATALYST.

NCI hits were subsequently filtered based on molecular weight, such that only hits of molecular weights ≤550 Da were retained. Surviving hits were either docked into CaNMT (PDB code: 1IYK) or GP (PDB code: 1L5R), employing the docking conditions of corresponding dbCICA models (**A-II** or **B−III** as in Tables 1 and 6, respectively). The resulting docked poses were subsequently analyzed for critical contacts according to dbCICA models **A-II** or **B−III**, and the sums of critical contacts for each hit compound were used to rank hits. The highest ranking available hits were acquired and tested in vitro.

4.2. Microbiological Assays. *4.2.1. Materials and Instruments.* Nutrient media of broth and agar (Sabouraud dextrose and Mueller-Hinton) were obtained from Oxoid Ltd. (England) and were reconstituted in accordance with the manufacturer's recommendations. Dimethylsulfoxide (DMSO) was purchased from Tedia Company (USA). Miconazole was used as reference antifungal agent and was obtained as an in kind gift from Dar Al-Dawa Manufacturing Company, Na'ur, Jordan. Hit compounds were obtained as an in kind gift from the National Cancer Institute (NCI) (Maryland, USA) and were tested without further purification. A total of 96 flat bottom microtiter plates (TPP Company, Switzerland), WTC binder static incubator (J.P. Selecta, Spain), Daigi shaking incubator (Daigi Sciences Co., Ltd., Korea), and a Raypa steam sterilizer (R. Espinar S.C, Spain) were used.

4.2.2. Microbial Strains. The antifungal and antibacterial activities of hit compounds were evaluated against standard reference *Candida albicans* (*C. albicans*) ATCC 10231, *Staphylococcus aureus* (*S. aureus*) ATCC 6538P, and *Escherichia coli* (*E. coli*) ATCC 8739 strain.

4.2.3. Broth Microdilution Method. The minimum inhibitory concentration (MIC) of the hit compounds against *C. albicans*, *S. aureus*, and *E. coli* was determined via broth microdilution method according to the Clinical and Laboratory Standards Institute.[83]

MIC test was performed in 96 flat bottom microtiter plates. Each hit compound was dissolved in minimum a mount of DMSO and used as a stock solution. The stock solutions were stored at 4 °C in tightly capped amber glass bottles. A suitable volume, 20 μL, was taken from the stock solution by means of micropipet and was added to the first column of test well, and then the volume was completed up to 200 μL using single strength Sabouraud dextrose broth for *C. albicans* (or Mueller-Hinton broth for *E. coli* and *S. aureus*). A series of 2-fold dilutions was then carried out across the plate, changing the tips at each dilution step. Then, 10 μL of preadjusted overnight microbial culture was used to inoculate each well in the microtiter plate to achieve a final inoculum size of about 1.15×10^6 cfu/well for *C. albicans*, about 1.75×10^6 cfu/well for *E. coli*, and about 3.72×10^6 cfu/well for *S. aureus*. Then the 96 microtiter plate was incubated for 24 h using the static incubator. All assays included positive and negative growth controls. Positive controls are compound-free wells (containing DMSO, overnight culture microbial inoculum, and nutrient medium), while negative growth controls are wells without either hit compound or DMSO (only overnight culture microbial inoculum and nutrient medium). MICs were expressed as the average of two successive concentrations of the antimicrobial agent showing no growth (clear) and growth (turbid). MIC determination was carried out in triplicate.

4.2.4. In Vitro GP Enzyme Inhibition Assay. GPa (from rabbit muscle), glycogen, glucose-1-phosphate, malachite green, and ammonium molybdate were purchased from the Sigma-Aldrich Corporation (St. Louis, MO, USA). Reagents and solvents were obtained from commercial suppliers and used without further purification.

The enzymatic inhibition of phosphorylase activity was monitored using a microplate reader (BioTek, USA) on the basis of published methods.[118] In brief, GPa activity was measured in the direction of glycogen synthesis by the release of phosphate from glucose-1-phosphate. Each compound was dissolved in DMSO and diluted appropriately. The enzyme was added into the 100 μL buffer containing 50 mM Hepes (pH 7.2), 100 mM KCl, 2.5 mM $MgCl_2$, 0.5 mM glucose-1-phosphate, and 1 mg/mL glycogen together with tested compounds in 96 well microplates. After the addition of 150 μL of 1 M HCl containing 10 mg/mL ammonium molybdate and 0.38 mg/mL malachite green, reactions were run at 22 °C for 20 min, and then the phosphate absorbance was measured at 620 nm. Caffeine was used as standard inhibitor for GP enzyme (IC_{50} = 120 μM).[111]

■ ASSOCIATED CONTENT

Supporting Information. Optimal conditions and parameters necessary for successful docking of 108 CaNMT inhibitors, contacts thresholds in Å, number of positive and negative contacts, and statistical criteria of the best dbCICA models based on docked solutions generated by FRED, LibDock, and LigandFit, *X*, *Y*, *Z* coordinates of the two pharmacophores. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: 00962-6-5355000 ext. 23305; fax: 00962-6-5339649; e-mail: mutasem@ju.edu.jo.

■ REFERENCES

(1) Song, C. M.; Lim, S. J.; Tong, J. C. Recent advances in computer-aided drug design. *Brief Bioinform.* **2009**, *10*, 579–591.

(2) Menikarachchi, L. C.; Gascon, J. A. QM/MM approaches in medicinal chemistry research. *Curr. Top. Med. Chem.* **2010**, *10*, 46–54.

(3) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42*, 724–733.

(4) Hecht, D.; Fogel, G. B. Computational intelligence methods for docking scores. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 56–68.
(5) Steinbrecher, T.; Labahn, A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* **2010**, *17*, 767–785.
(6) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein−ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
(7) Taha, M. O.; AlDhamin, M. Effects of variable docking conditions and scoring functions on the qualities of protein aligned CoMFA models constructed from diverse h-PTP 1B inhibitors. *J. Med. Chem.* **2005**, *48*, 8016–8034.
(8) Krovat, E. M.; Langer, T. Impact of scoring functions on enrichment in docking-based virtual screening: An application study on renin inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1123–1129.
(9) Beeley, N. R. A.; Sage, C. GPCRs: An update on structural approaches to drug discovery. *Targets* **2003**, *2*, 19–25.
(10) Klebe, G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.
(11) Steuber, H.; Zentgraf, M.; Gerlach, C.; Sotriffer, C. A.; Heine, A; Klebe, G. Expect the unexpected or caveat for drug designers: Multiple structure determinations using aldose reductase crystals treated under varying conditions. *J. Mol. Biol.* **2006**, *363*, 174–187.
(12) Stubbs, M. T.; Reyda, S.; Dullweber, F.; Moller, M.; Klebe, G.; Dorsch, D.; Mederski, W.; Wurziger, H. pH-dependent binding modes observed in trypsin crystals: Lessons for structure-based drug design. *ChemBioChem* **2002**, *3*, 246–249.
(13) DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004**, *12*, 831–838.
(14) Morris, G. M.; Olson, A. J.; Goodsell, D. S. Protein−ligand docking methods. *Princ. Med. Chem.* **2000**, *8*, 31–48.
(15) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558–565.
(16) Beier, C.; Zacharias, M. Tackling the challenges posed by target flexibility in drug design. *Expert Opin. Drug Dis.* **2010**, *5*, 347–359.
(17) Boyd, S. FlexX suite. *Chem. World* **2007**, *4*, 72.
(18) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(19) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aid Mol. Des.* **2001**, *15*, 411–428.
(20) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(21) Vaque, M.; Ardrevol, A.; Blade, C.; Salvado, M. J; Blay, M.; Fernandez-Larrea, J.; Arola, L.; Pujadas, G. Protein−ligand docking: A review of recent advances and future perspectives. *Curr. Pharm. Anal.* **2008**, *4*, 1–19.
(22) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual screening with AutoDock: Theory and practice. *Expert Opin. Drug Dis.* **2010**, *5*, 597–607.
(23) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(24) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
(25) *CERIUS2 LigandFit*, version 4.10; Accelrys, Inc.: San Diego, 2000.
(26) *FRED*, version 2.1; OpenEye Scientific Software: Santa Fe, NM, 2006.
(27) Diller, D. J.; Merz, K. M. High throughput docking for library design and library prioritization. *Proteins* **2001**, *43*, 113–124.
(28) Rao, S. N.; Head, M. S.; Kulkarni, A.; LaLonde, J. M. Validation studies of the site-directed docking program LibDock. *J. Chem. Inf. Model.* **2007**, *47*, 2159–2171.
(29) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
(30) Gao, W. R.; Lai, Y. L. SCORE: A new empirical method for estimating the binding affinity of a protein−ligand complex. *J. Mol. Model.* **1998**, *4*, 379–394.
(31) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395–407.
(32) Velec, H. F. G.; Gohlke, H.; Klebe, G. Drug score knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
(33) Jain, A. N. Scoring functions for protein−ligand docking. *Curr. Protein Pept. Sci.* **2006**, *7*, 407–20.
(34) Rajamani, R.; Good, A. C. Ranking poses in structure-based lead discovery and optimization: Current trends in scoring function development. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 308–15.
(35) Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, *13*, 3583–608.
(36) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J. Chem. Inf. Model.* **2009**, *49*, 1568–1580.
(37) Jain, A. N. Scoring non-covalent protein−ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 427–440.
(38) Böhm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3-D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309–323.
(39) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.
(40) Wang, R.; Gao, Y.; Lai, L. SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *J. Mol. Model.* **1998**, *4*, 379–394.
(41) Gehlhaar, D. K.; Bouzida, D.; Rejto, P. Reduced dimensionality in ligand-protein structure prediction: Covalent inhibitors of serine proteases and design of site-directed combinatorial libraries. In *Rational Drug Design: Novel Methodology and Practical Applications*; Parrill, L.; Rami Reddy, M.; Eds.; ACS Symposium Series 719; American Chemical Society; Washington, DC, 1999; pp 292−311.
(42) Wang, R.; Lai, L.; Wang, S. Further development and of empirical scoring functions for structure-based binding validation affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.
(43) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.
(44) Muegge, I. A knowledge-based scoring function for protein−ligand interactions: Probing the reference state. *Perspect. Drug Discov.* **2000**, *20*, 99–114.
(45) Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem.* **2001**, *22*, 418–425.
(46) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
(47) Muegge, I. PMF scoring revisited. *J. Med. Chem.* **2006**, *49*, 5895–902.
(48) Tame, J. R. H. Scoring functions: A view from the bench. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 99–108.
(49) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(50) Homans, S. W. Water, water everywhere—except where it matters. *Drug Discovery Today* **2007**, *12*, 534–539.
(51) Poornima, C. S.; Dean, P. M. Hydration in drug design. 1. Multiple hydrogen- bonding features of water molecules in mediating protein−ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500–512.
(52) Poornima, C. S.; Dean, P. M. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 513–520.
(53) Poornima, C. S.; Dean, P. M. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521–531.
(54) Koehler, K. F.; Rao, S. N.; Snyder, J. P. Modeling drug-receptor interactions. In *Guidebook on Molecular Modeling in Drug Design*; Cohen, N. C., Ed.; Academic Press; San Diego, 1996; pp 235−336.
(55) Pastor, M.; Cruciani, G.; Watson, K. A Strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure−activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089–4102.
(56) Silverman, R. A. The Organic Chemistry of Drug Design and Drug Action; Academic Press: San Diego, 1991, pp 62−65.
(57) Sutherland, J. J.; Nandigam, R. K.; Erickson, J. A.; Vieth, M. Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. *J. Chem. Inf. Model.* **2007**, *47*, 2293–2302.
(58) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
(59) Abu-Hammad, A. M.; Afifi, F.; Taha, M. O. Combining docking, scoring, and molecular field analyses to probe influenza neuraminidase-ligand interactions. *J. Mol. Graphics Modell.* **2007**, *26*, 443–456.
(60) Abu-Hammad, A.; Zalloum, W. A.; Zalloum, H.; Abu-Sheikha, G.; Taha, M. O. Homology modeling of MCH1 receptor and validation by docking/scoring and protein-aligned CoMFA. *Eur. J. Med. Chem.* **2009**, *44*, 2583–2596.
(61) *Discovery Studio*, version 2.5; Accelrys, Inc.: San Diego, 2009.
(62) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. application to high-throughput docking on metabotropic glutamate receptor subtype. *J. Med. Chem.* **2005**, *48*, 2534–2547.
(63) Triballeau, N.; Bertrand, H. O.; Acher, F. Are You Sure You Have a Good Model? In *Pharmacophores and Pharmacophore Searches*; Langer, T., Hoffmann, R. D., Eds.; Wiley-VCH, Verlag GmbH & Co.: Weinheim, Germany, 2006; pp 325−364.
(64) Masubuchi, M.; Kawasaki, K.; Ebiike, H.; Ikeda, Y.; Tsujii, S.; Sogabe, S.; Fujii, T.; Sakata, K.; Shiratori, Y.; Aoki, Y.; Ohtsuka, T.; Shimma, N. Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 1. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1833–1837.
(65) Ebiike, H.; Masubuchi, M.; Liu, P.; Kawasaki, K.; Morikami, K.; Sogabe, S.; Hayase, M.; Fujii, T.; Sakata, K.; Shindoh, H.; Shiratori, Y.; Aoki, Y.; Ohtsukaa, T.; Shimma, N. Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 2. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 607–610.
(66) Kawasaki, K.; Masubuchi, M.; Morikami, K.; Sogabe, S.; Aoyama, T.; Ebiike, H.; Niizuma, S.; Hayase, M.; Fujii, T.; Sakata, K.; Shindoh, H.; Shiratori, Y.; Aoki, Y.; Ohtsuka, T.; Shimma, N. Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 3. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 87–91.
(67) Masubuchi, M.; Ebiike, H.; Kawasaki, K.; Sogabe, S.; Morikami, K.; Shiratori, Y.; Tsujii, S.; Fujii, T.; Sakata, K.; Hayase, M.; Shindoh, H.; Aoki, Y.; Ohtsuka, T.; Shimma, N. Synthesis and biological activities of benzofuran antifungal agents targeting fungal N-myristoyltransferase. *Bioorg. Med. Chem.* **2003**, *11*, 4463–4478.
(68) Yamazaki, K.; Kaneko, Y.; Suwa, K.; Ebara, S.; Nakazawa, K.; Yasuno, K. Synthesis of potent and selective inhibitors of *Candida albicans* N-myristoyltransferase based on the benzothiazole structure. *Bioorg. Med. Chem.* **2005**, *13*, 2509–2522.
(69) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
(70) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.
(71) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 Protease: Conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317–324.
(72) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.
(73) Verkivker, G. M.; Bouzida, D.; Gehlaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Deciphering common failures in molecular docking of ligand−protein complexes. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 731–751.
(74) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
(75) Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.
(76) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035–1042.
(77) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson−Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
(78) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Watson, P. Virtual screening using protein−ligand docking: Avoiding artificial enrichment. *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 793–806.
(79) Irwin, J. J.; Shoichet, B. K. ZINC: A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(80) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3-D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection . What can we learn from earlier mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213–228.
(81) Jacobsson, M.; Liden, P.; Stjernschantz, E.; Bostroem, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.* **2003**, *46*, 5781–5789.
(82) Gao, H.; Williams, C.; Labute, P.; Bajorath, J. Binary quantitative structure−activity relationship (QSAR) analysis of estrogen receptor ligands. *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 164–168.
(83) Performance Standards for Antimicrobial Susceptibility Testing, 15th Information Supplement; Document M100-S15; Clinical and Laboratory Standards Institute (CLSI): Wayne, PA, 2005.
(84) Waszkowycz, B. New Methods for Structure-Based De Novo Drug Design. In *Advances in Drug Discovery Techniques*; Harvey, A. L., Ed.; John Wiley & Sons: West Sussex, U.K., 1998; pp 150−153.
(85) Krissinel, E. Crystal contacts as nature's docking solutions. *J. Comput. Chem.* **2009**, *31*, 133–143.
(86) Mcwherter, C. A.; Rocque, W. J.; Wood, D. C.; Gordon, J. I. A comparative analysis of the kinetic mechanism and peptide substrate specificity of human and *Saccharomyces cerevisiae* myristoyl-CoA:protein N-myristoyltransferase. *J. Biol. Chem.* **1993**, *268*, 9964–9971.
(87) Gordon, J. I.; Duronio, R. J.; Rudnick, D. A.; Adams, S. P.; Gokel, G. W. Protein N-myristoylation. *J. Biol. Chem.* **1991**, *266*, 8647–8650.
(88) Farazi, T. A.; Waksman, G.; Gordon, J. I. The biology and enzymology of protein N-myristoylation. *J. Biol. Chem.* **2001**, *276*, 39501–39504.

(89) Weston, S. A.; Camble, R.; Colls, J.; Rosenbrock, G.; Taylor, I.; Egerton, M.; Tucker, A. D.; Tunnicliffe, A.; Mistry, A.; Mancia, F. Crystal structure of the anti-fungal target N-myristoyl transferase. *Nat. Struct. Biol.* **1998**, *5*, 213–221.

(90) Sogabe, S.; Masubuchi, M.; Sakata, K.; Fukami, T. A.; Morikami, K.; Shiratori, Y.; Ebiike, H.; Kawasaki, K.; Aoki, Y.; Shimma, N. Crystal structures of *Candida albicans* N-myristoyltransferase with two distinct inhibitors. *Chem. Biol.* **2002**, *9*, 1119–1128.

(91) Wu, J.; Tao, Y.; Zhang, M.; Howard, M. H.; Gutteridge, S.; Ding, J. Crystal structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myristoyl-CoA and inhibitors reveal the functional roles of the N-terminal region. *J. Biol. Chem.* **2007**, *282*, 22185–22194.

(92) Ramsey, L. F.; Schafer, W. D. *The Statistical Sleuth*, 1st ed.; Wadsworth Publishing Company: Belmont, CA, 1997.

(93) Ercan-Fang, N.; Taylor, R. M.; Treadway, L. J.; Levy, B. C.; Genereux, E. P.; Gibbs, M. E.; Rath, L. V.; Kwon, Y.; Gannon, C. M.; Nuttall, Q. F. Endogenous effectors of human liver glycogen phosphorylase modulate effects of indole-site inhibitors. *Am. J. Physiol. Endocrinol. Metab.* **2005**, *289*, E366–E372.

(94) Tsirkone, V. G.; Tsoukala, E.; Lamprakis, C.; Manta, S.; Hayes, J. M.; Skamnaki, V. T.; Drakou, C.; Zographos, S. Y.; Komiotis, D.; Leonidas, D. D. 1-(3-Deoxy-3-fluoro-b-D-glucopyranosyl) pyrimidine derivatives as inhibitors of glycogen phosphorylase b: Kinetic, crystallographic, and modeling studies. *Bioorg. Med. Chem.* **2010**, *18*, 3413–25.

(95) Zhang, L.; Li, H.; Zhu, Q.; Liu, J.; Chen, L.; Leng, Y.; Jiang, H.; Liu, H. Benzamide derivatives as dual-action hypoglycemic agents that inhibit glycogen phosphorylase and activate glucokinase. *Bioorg. Med. Chem.* **2009**, *13*, 4385–4388.

(96) Thomson, S. A.; Banker, P.; Bickett, D. M.; Boucheron, J. A.; Carter, H. L.; Clancy, D. C.; Cooper, J. P.; Dickerson, S. H.; Garrido, D. M.; Nolte, R. T.; Peat, A. J.; Sheckler, L. R.; Sparks, S. M.; Tavares, F. X.; Wang, L.; Wang, T. Y.; Weiel, J. E. Anthranilimide based glycogen phosphorylase inhibitors for the treatment of type 2 diabetes. Part 3: X-ray crystallographic characterization, core and urea optimization and in vivo efficacy. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1177–1182.

(97) Onda, K.; Suzuki, T.; Shiraki, R.; Yonetoku, Y.; Negoro, K.; Momose, K.; Katayama, N.; Orita, M.; Yamaguchi, T.; Ohta, M.; Tsukamoto, S. Synthesis of 5-chloro-N-aryl-1H-indole-2-carboxamide derivatives as inhibitors of human liver glycogen phosphorylase a. *Bioorg. Med .Chem.* **2008**, *16*, 5452–5464.

(98) Anderka, O.; Loenze, P.; Klabunde, T.; Dreyer, M. K.; Defossa, E.; Wendt, K. U.; Schmoll, D. Thermodynamic characterization of allosteric glycogen phosphorylase inhibitors. *Biochemistry* **2008**, *47*, 4683–4691.

(99) Pautsch, A.; Stadler, N.; Wissdorf, O.; Langkopf, E.; Moreth, W.; Streicher, R. Molecular recognition of the protein phosphatase 1 glycogen targeting subunit by glycogen phosphorylase. *J. Biol. Chem.* **2008**, *283*, 8913–8918.

(100) Klabunde, T.; Wendt, K. U.; Kadereit, D.; Brachvogel, V.; Burger, H. J.; Herling, A. W.; Oikonomakos, N. G.; Kosmopoulou, M. N.; Schmoll, D.; Sarubbi, E.; von Roedern, E.; Defossa, E. Acyl ureas as human liver glycogen phosphorylase inhibitors for the treatment of type 2 diabetes. *J. Med. Chem.* **2005**, *48*, 6178–6193.

(101) Hamada, T.; Ito, Y.; Abe, T.; Hayashi, F.; Guntert, P.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yoshida, M.; Tanaka, A.; Sugano, S.; Yokoyama, S.; Hirota, H. Solution structure of the antifreeze-like domain of human sialic acid synthase. *Protein Sci.* **2006**, *15*, 1010–1016.

(102) Rath, V. L.; Ammirati, M.; Danley, D. E.; Ekstrom, J. L.; Gibbs, E. M.; Hynes, T. R.; Mathiowetz, A. M.; McPherson, R. K.; Olson, T. V.; Treadway, J. L.; Hoover, D. J. Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chem. Biol.* **2000**, *7*, 677–682.

(103) Ekstrom, J. L.; Pauly, T. A.; Carty, M. D.; Soeller, W. C.; Culp, J.; Danley, D. E.; Hoover, D. J.; Treadway, J. L.; Gibbs, E. M.; Fletterick, R. J.; Day, Y. S.; Myszka, D. G.; Rath, V. L. Structure−activity analysis of the purine binding site of human liver glycogen phosphorylase. *Chem. Biol.* **2002**, *9*, 915–924.

(104) Oikonomakos, N. G.; Kosmopoulou, M. N.; Chrysina, E. D.; Leonidas, D. D.; Kostas, I. D.; Wendt, K. U.; Klabunde, T.; Defossa, E. Crystallographic studies on acyl ureas, a new class of glycogen phosphorylase inhibitors, as potential antidiabetic drugs. *Protein Sci.* **2005**, *14*, 1760–1771.

(105) Wright, S. W.; Rath, V. L.; Genereux, P. E.; Hageman, D. L.; Levy, C. B.; McClure, L. D.; McCoid, S. C.; McPherson, R. K.; Schelhorn, T. M.; Wilder, D. E.; Zavadoski, W. J.; Gibbs, E. M.; Treadway, J. L. 5-Chloroindoloyl glycine amide inhibitors of glycogen phosphorylase: Synthesis, in vitro, in vivo, and X-ray crystallographic characterization. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 459–465.

(106) Patera, A.; Blaszczak, L. C.; Shoichet, B. K. Crystal structures of substrate and inhibitor complexes with AmpC−Lactamase: Possible implications for substrate-assisted catalysis. *J. Am. Chem. Soc.* **2000**, *122*, 10504–10512.

(107) Rath, V. L.; Ammirati, M.; LeMotte, P. K.; Fennell, K. F.; Mansour, M. N.; Danley, D. E.; Hynes, T. R.; Schulte, G. K.; Wasilko, D. J.; Pandit, J. Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core. *Mol. Cell* **2000**, *6*, 139–148.

(108) Zographos, S. E.; Oikonomakos, N. G.; Tsitsanou, K. E.; Leonidas, D. D.; Chrysina, E. D.; Skamnaki, V. T.; Bischoff, H.; Goldmann, S.; Watson, K. A.; Johnson, L. N. The structure of glycogen phosphorylase b with an alkyldihydropyridine-dicarboxylic acid compound, a novel and potent inhibitor. *Structure* **1997**, *5*, 1413–1425.

(109) Deng, Q.; Lu, Z; Bohn, J.; Ellsworth, K. P.; Myers, R. W.; Geissler, W. M.; Harris, G.; Willoughby, C. A.; Chapman, K.; McKeever, B.; Mosley, R. Modeling aided design of potent glycogen phosphorylase inhibitors. *J. Mol. Graph. Model.* **2005**, *23*, 457–464.

(110) Ekstrom, J. L.; Pauly, T. A.; Carty, M. D.; Soeller, W. C.; Culp, J.; Danley, D. E.; Hoover, D. J.; Treadway, J. L.; Gibbs, E. M.; Fletterick, R. J.; Day, Y. S.; Myszka, D. J.; Rath, V. L. Structure−activity analysis of the purine binding site of human liver glycogen phosphorylase. *Chem. Biol.* **2002**, *9*, 915–924.

(111) Loughlin, A. W.; Pierens, K. G.; Petersson, J. M.; Hendersona, C. L.; Healya, C. P. Evaluation of novel hyphodermin derivatives as glycogen phosphorylase a inhibitors. *Bioorg. Med. Chem.* **2008**, *16*, 6172–6178.

(112) Al-masri, I. M.; Mohammad, M. K.; Taha, M. O. Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. *ChemMedChem.* **2008**, *3*, 1763–1779.

(113) Taha, M. O.; Bustanji, Y.; Al-Bakri, A. G.; Yousef, A.-M.; Zalloum, W. A.; Al-Masri, I. M.; Atallah, N. Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in-silico screening. *J. Mol. Graph. Modell.* **2007**, *25*, 870–884.

(114) Al-Sha'er, M. A.; Taha, M. O. Discovery of novel CDK1 inhibitors by combining pharmacophore modeling, QSAR analysis and in silico screening followed by in vitro bioassay. *Eur. J. Med. Chem.* **2010**, *45*, 4316–4330.

(115) Al-Sha'er, M. A.; Taha, M. O. Elaborate ligand-based modeling reveal new nanomolar heat shock protein 90a inhibitors. *J. Chem. Info. Model.* **2010**, *50*, 1706–1723.

(116) Birch, A. M.; Kenny, P. W.; Oikonomakos, N. G.; Otterbein, L.; Schofield, P.; Paul, R. O.; Whittamore; Whalley, D. P. Development of potent, orally active 1-substituted-3,4-dihydro-2-quinolone glycogen phosphorylase inhibitors. *Bioorg. Med. Chem.* **2007**, *17*, 394–399.

(117) Rosauer, K. G.; Ogawa, A. K.; Willoughby, C. A.; Ellsworth, K. P.; Geissler, W. M.; Myers, W. R.; Chapman, K. T.; Harris, G.; Moller, D. E. Novel 3,4-Dihydroquinolin-2(1H)-one Inhibitors of human glycogen phosphorylase a. *Bioorg. Med. Chem.* **2003**, *13*, 4385–4388.

(118) Martin, W. H.; Hoover, D. J.; Armento, S. J.; Stock, I. A.; Mcpherson, R. K.; Danley, D. E.; Stevenson, R. W.; Barrett, E. J.; Tredway, J. L. Discovery of a human liver glycogen phosphorylase inhibitor that lowers blood glucose in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1776–1781.